This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

B 1,385,073

THE UNIVERSITY OF MICHIGAN
LIBRARIES
THE UNIVERSITY OF MICHIGAN
1817

PRINCIPLES OF PHYSICS

DESIGNED FOR USE AS A TEXTBOOK
OF GENERAL PHYSICS

BY

WILLIAM FRANCIS MAGIE, PH.D.
HENRY PROFESSOR OF PHYSICS IN PRINCETON UNIVERSITY

NEW YORK
THE CENTURY CO.
1911

hysics
QC
21
.M18

COPYRIGHT, 1911,
BY
THE CENTURY CO.

Published, May, 1911

Physics Lit.
Stanley Lit
5-2-39

PREFACE

THE subject of Physics forms a part of almost every curriculum of liberal studies. It holds this place not only because of the intrinsic interest of its subject-matter, but also because of the way in which it exemplifies the scientific method of thinking. In this latter respect it is unrivaled. The observations upon which a physical law is based are in most cases few in number, and the application of the inductive process to them is immediate and obvious. Furthermore, the phenomena observed are of so many different sorts, and the laws which are obtained from them are so numerous, that in dealing with them the inductive process is applied again and again. When compared with the other sciences, physics is the one in which scientific reasoning is illustrated by the simplest and most varied examples. For these reasons it is peculiarly adapted to train the mind in the method of thought which is used in scientific study and investigation and, indeed, in the practical business of life.

It has for some time been my belief that, as physics is now commonly taught, these peculiar advantages of the science as a means of intellectual training are not given sufficient attention.

Urged by this belief I have endeavored in this book to present the subject in such a way as to direct attention particularly to the development of its various branches and thus to exemplify the processes of thought which are employed in the examination of a group of physical phenomena and the establishment of a physical law or theory. A general guide to the attainment of this end is found in the history of physics, and accordingly I have constructed the exposition of the subject upon the historical outline.

Besides the assistance which it furnishes in this most important respect the historical order has certain peculiar advantages. It puts in the clearest light those facts of the science which are fundamental, and shows the way in which theory grows out of facts. It leads the mind from topic to topic along the path of discovery, so as to arouse its curiosity and at the same time to direct its attention to

the reasons which render each step taken a natural one to take. It also presents the subject in the way in which it can most easily be apprehended. The progress of discovery has been along the lines of least intellectual resistance, and it is probable that what was easiest to discover once will now be the easiest to understand. It further enables us to pursue the argument whereby a physical law is established, sometimes almost in the words, and generally along the course of thought, of its original discoverer, so that the mind is brought into intimate relationship with the great minds of the past, and experiences the intellectual exhilaration which such a relationship always brings with it.

A fairly faithful adherence to my plan has resulted in the necessary omission of many details of fact and theory which are commonly presented in textbooks of physics. For these the student must look to the course of experimental lectures or of laboratory practice, one or both of which should form part of a course of instruction in which this book is used.

To illustrate the principles discussed and to introduce many important theorems without breaking up too much the continuity of the argument, I have made use of the examples which are placed in groups in various parts of the book. With the same purpose, some of the demonstrations of theorems which are of first importance in themselves, but the formal demonstration of which is of secondary importance for the student, have been put in fine print.

The use of the historical outline accounts for what is nowadays the unusual order followed in the presentation of some of the subjects. It leads to a much fuller discussion of statics than is generally given, and to other peculiar features of the treatment, such as the use of the statical measure of force, the belated introduction of the relations of heat to energy, the use of the method of rays in geometrical optics, and to other minor divergences from common practice. In the alternative methods which are frequently employed, the true inductive order is either avoided or obscured, and the course of thought followed is neither natural nor that which would be followed by an investigator.

I have ventured to name the book Principles of Physics, not to make a claim that it presents a complete and critical study of the foundations of the subject, but to indicate the object which, however imperfectly, I have tried to attain, that is, the exposition of the principal physical laws and theories in the light of the arguments by which they were established.

My friends Professor Howard McClenahan and Professor Augustus Trowbridge have given me great assistance by their examination and criticism of the manuscript. I am indebted to Mr. E. S. Taylerson, Research Assistant in the Palmer Physical Laboratory, for his help in preparing the diagrams.

W. F. MAGIE.

PRINCETON UNIVERSITY,
February, 1911.

CONTENTS

PAGE

INTRODUCTION, §§ 1–10 3

BOOK I

MECHANICS

CHAPTER

I. GENERAL NOTIONS ABOUT FORCES, §§ 11–15 9

II. STATICS, §§ 16–48 12

Principle of Moments, §§ 17–30; Examples, I 12

Principle of Work, §§ 31–37; Examples, II 29

The Parallelogram of Forces, §§ 38–45; Examples, III 38

Static Friction, §§ 46–48; Examples, IV 49

III. DYNAMICS, §§ 49–116 53

Dynamics of a Particle, §§ 50–88 53

Force and Motion, §§ 50–66; Examples, V 53

Mass and Momentum, §§ 67–79; Examples, VI 77

Work and Energy, §§ 80–88; Examples, VII 88

Dynamics of Extended Bodies, §§ 89–116 96

Center of Mass, §§ 89–93; Examples, VIII 96

Uniplanar Motion of Rigid Bodies, §§ 94–107; Examples, IX 103

General Motion of Rigid Bodies, §§ 108–112 120

Kinetic Friction, §§ 113–116; Examples, X 126

IV. MECHANICS OF LIQUIDS, §§ 117–135 130

Hydrostatics, §§ 117–128 130

Hydrodynamics, §§ 129–135 141

BOOK II

GENERAL PROPERTIES OF MATTER

V. GRAVITATION OR MASS ATTRACTION, §§ 136–148; Examples, XI 149

VI. ELASTICITY, §§ 149–164; Examples, XII 166

VII. CAPILLARITY, §§ 165–176; Examples, XIII 176

VIII. GASES, §§ 177–188 188

IX. FRICTION, §§ 189–195 202

X. DIFFUSION, §§ 196–200 206

BOOK III

SOUND

Chapter		Page
XI.	Sound and Sound Waves, §§ 201–213; Examples, XIV	211
XII.	Sounding Bodies, §§ 214–227	229

BOOK IV

HEAT

XIII.	Temperature and Heat, §§ 228–256	245
	Heat, §§ 241–253	252
	Transfer of Heat, §§ 254–256	262
XIV.	Thermal Relations of Bodies, §§ 257–276	267
XV.	Heat and Energy, §§ 277–295; Examples, XV	284

BOOK V

LIGHT

XVI.	Geometrical Optics, §§ 296–335	309
	Reflection of Light, Mirrors, §§ 304–307; Examples, XVI	312
	Refraction of Light, Lenses, §§ 308–320; Examples, XVII	319
	Optical Instruments, §§ 321–327; Examples, XVIII	339
	Dispersion, §§ 328–330	346
	Velocity, §§ 331–335	348
XVII.	Physical Optics, §§ 336–376	352
	Wave Theory, §§ 341–345	355
	Interference and Diffraction, §§ 346–355; Examples, XIX	361
	Polarization, §§ 356–363	376
	The Spectrum, §§ 364–376	386

BOOK VI

MAGNETISM AND ELECTRICITY

XVIII.	Magnetism, §§ 377–409; Examples, XX	395
	Magnetic Field, §§ 390–400; Examples, XXI	406
	Magnetic Induction, §§ 401–409; Examples, XXII	418
XIX.	Static Electricity, §§ 410–449; Examples, XXIII	429
	Electric Field and Electric Induction, §§ 437–449; Examples, XXIV	446

CHAPTER PAGE

XX. THE ELECTRIC CURRENT, §§ 450–507 461
Electromagnetism, §§ 458–479; Examples, XXV 466
Electrochemistry, §§ 480–490 487
Thermoelectricity, §§ 491–492 496
General Laws of the Electric Current, §§ 493-507; Examples, XXVI; Examples, XXVII 497

XXI. ELECTRICITY AND THE ETHER, §§ 508–513 520

XXII. ELECTRICITY AND MATTER, §§ 514–525 526
The Electric Discharge, §§ 514–522 526
Radioactivity, §§ 523–525 535

EXERCISES 541

TABLES 557

INDEX 563

PRINCIPLES OF PHYSICS

PRINCIPLES OF PHYSICS

INTRODUCTION

1. The Science of Physics. — The word *physics* was used originally to designate that which pertains to the order of external nature, and therefore the science of external nature. In process of time and with the growth of knowledge, this universal science has been divided into many branches, discriminated from each other by their subject-matter and their method. The one of these to which the name *physics* is now applied deals particularly with what used to be called the physical agents, or with what we may now call, without speaking philosophically, the causes of our sense perceptions. Thus the branch of physics called mechanics originates in the sensation of force, acoustics in that of sound, heat in that of temperature, optics in that of light. Electricity and magnetism, to which we have no special sense to correspond, originate in our perception of the forces which are exhibited by electrified bodies or magnets. Each of these branches of physics has been developed from the first knowledge given by our sensations into a body of laws descriptive of the operations of natural bodies. It is generally the case in physics, in distinction from the other sciences, that these laws are capable of expression in mathematical formulas.

Because of the comparative simplicity of the physical laws, as also because of their fundamental character and general applicability, the science of physics used to be called, and is still sometimes called, natural philosophy.

2. The Scientific Method. — The intellectual process which is followed in the study of science in general is called induction. In the particular form of the process of induction used in the study of physics, the first step is the gathering of material by the *observation* of natural phenomena. Our observations are sometimes

made on phenomena occurring spontaneously in nature and not subject to our control; but ordinarily they are obtained by *experiments* upon natural bodies, in which we arrange the bodies in such a way as to suppress, as far as possible, those phenomena with which we are not immediately concerned, and to make prominent those which we desire to observe. The results of these experiments are considered and an attempt is made to discover, in the phenomena presented by them, some simple general relation from which the existence of a rule or *law* of procedure may be inferred. When the law is once perceived, additional observations of the same sort as those which led to its discovery may be made to confirm it, by multiplying examples of its applicability.

A further and more severe test of the suspected law is made by using it as a postulate or *hypothesis* from which conclusions are drawn by the methods of deductive logic, which often take the form of mathematical operations, and interpreting these conclusions as descriptions of results which are necessarily true in fact if the law is valid. Experiments are then instituted in order to see whether the results thus predicted have their counterparts in nature.

If it is found that properly arranged experiments exhibit phenomena which are necessary results of the hypothesis, and no phenomena which are inconsistent with the hypothesis, our belief becomes very strong that the hypothesis which we have provisionally adopted is a valid description of a mode of procedure in nature, and we adopt it as expressing a *physical law*.

A physical law, as the phrase is generally used, describes or explains a limited group of phenomena. In order to construct a more general law, or *physical theory*, the various physical laws descriptive of a set of related phenomena are examined, and a hypothesis is made of some general mode of action of which we think it possible that the observed laws may be consequences. This hypothesis is tested, as in the process already described, by deducing from it its logical consequences. If these consequences agree with the physical laws already established, and if no consequences follow which do not agree with those laws, our confidence in the usefulness of our hypothesis as affording a valid description of phenomena is strengthened, and in some cases we attain a belief that the hypothesis really represents a mode of procedure in nature. In any case we call the hypothesis thus confirmed a physical theory.

3. **Measurement.** — In order that the physical phenomena which are observed can be connected by a law, especially if this law is to be expressed in mathematical form, they must be measured. *Measurement* consists in the comparison of the quantity to be measured with a unit of its own kind. The result of the comparison is a number which tells how many times the unit is contained in the quantity to be measured. In stating the result of the measurement this number must be given and the unit of measurement must be named.

4. **Physical Units.** — Strictly speaking, all physical units are chosen arbitrarily, either directly by the observer who is making the measurement, or by his consent to adopt a unit already chosen by somebody else. It is so important that the results of different observers shall be immediately comparable with each other, that conventional units have been adopted by common consent for almost all quantities which occur in physical investigations, and a conventional method agreed upon for constructing other units when they are needed.

Within the limited range of mechanics, it is found possible to construct a complete system of units by the arbitrary choice of three units, and the derivation of the others from those three by definition. The units usually chosen as *fundamental* in this system are those of mass, length, and time. The units obtained from these by definition are called *derived* units.

In the other parts of physics some physical law may always be found connecting some one of the quantities under observation directly by space or time relations with mechanical force or with energy, and thus a point of entry is established for the introduction of a system of derived units based upon the same three fundamental ones, supplemented when necessary by an additional or supplementary fundamental unit. Any unit determined in this way is called an *absolute unit*, and a system of such units is an *absolute system*.

The arbitrary element in the establishment of any derived unit is found in the definition by which it is made to depend upon the fundamental units.

5. **Unit of Mass.** — The unit of mass may be the ordinary unit, such as the pound or the kilogram, which is used in the common transactions of business. These units or standards are established, for the public welfare, by the authority of government.

The *pound* avoirdupois is a mass defined for England by act of

Parliament as the mass of a certain cylindrical block of platinum, preserved in the Standards Office at Westminster.

The *kilogram* is a mass agreed upon by the principal nations as the mass of a certain block of platinum-iridium which is in charge of the International Bureau of Weights and Measures, at Sèvres. The kilogram is equal very nearly to 2.2 pounds.

For most physical and chemical investigations these units are inconveniently large, and by common consent the gram, or the one one-thousandth part of the kilogram, is used as the unit of mass in physical measurement, and is adopted as the fundamental unit of mass in the construction of the absolute system of units.

6. Unit of Length. — The unit of length may be the ordinary unit, such as the yard or the meter, which is used in the common transactions of business. As in the case of the unit of mass, these units or standards are established, for the public welfare, by the authority of government.

The *yard* is a length defined for England by act of Parliament as the distance between two lines drawn on gold plugs inserted into a certain gun-metal bar which is preserved in the Standards Office at Westminster. The yard in the United States is defined by act of Congress as the 36/39.37 of a meter.

The *meter* is a length agreed upon by the principal nations as the distance between two lines drawn on a bar of platinum-iridium which is in charge of the International Bureau of Weights and Measures at Sèvres.

For most physical investigations these units are inconveniently large, and by common consent the centimeter, or the one one-hundredth part of a meter, is used as the unit of length in physical measurement, and is adopted as the fundamental unit of length in the construction of the absolute system of units.

7. Unit of Time. — The time interval which serves in common life as the basis of measurement of time is the time of rotation of the earth with respect to the sun. This time is called the day. The average length of the day for a year is the *mean solar day*, which can be accurately determined, and serves as the basis of the scientific measurement of time.

In physical measurements, the unit of time adopted by common consent is the second, or the 1/86400 of the mean solar day. This unit is used in most physical measurements, and is adopted as the fundamental unit of time in the construction of the absolute system of units.

8. Unit of Angle. — The measurement of angle is so frequent in physical operations, and the use of angles so common in physical formulas, that it is important to call attention to the fact that in all our theoretical formulas, unless the contrary is explicitly stated, angles are measured in radians. The unit angle, called the *radian*, is the angle between the radii of a circle which subtends an arc of the circle equal to the radius. Any other angle is then measured by the ratio of the arc which it subtends to the radius of that arc.

With the radian as the unit, a right angle is equal to $\pi/2$. Measured in degrees the radian equals 57.2957795.

9. The c.g.s. System. — The system of absolute units, in which the centimeter is the unit of length, the gram the unit of mass, and the second the unit of time, is called the *centimeter-gram-second system*, and is commonly referred to as the *c.g.s. system* of units. It is the one which will be gradually developed in the sequel.

10. Dimensions of Physical Quantities. — In giving the measure of any physical quantity in c.g.s. units it is often convenient to follow the number which expresses the ratio of the quantity to its unit with a group of symbols representing the gram, the centimeter, and the second in the relation in which they enter in the definition of the unit. Thus a velocity of 10 units may be written 10 cm/sec; an acceleration of 10 units, 10 cm/sec^2; a force of 10 units, 10 gr cm/sec^2; an energy of 10 units, 10 gr cm^2/sec^2. This mode of expressing the unit, which may be extended to all parts of the absolute system, represents in an explicit manner the way in which the fundamental units enter into the definition of the derived unit.

In a similar but more general way we may use combinations of the symbols M for mass, L for length, and T for time, to represent the way in which the fundamental concepts of mass, length, and time enter into the definition of the derived concept. Such a combination of symbols is called the *dimensional formula* of the physical concept. The exponents of the symbols in this formula are called the *dimensions* of the concept in mass, length, and time respectively. Thus the dimensional formula of velocity is LT^{-1}, of acceleration LT^{-2}, of force MLT^{-2}, of energy ML^2T^{-2}. Every physical concept which is defined by its relation, either immediate or derived, to the fundamental concepts, has its appropriate dimensional formula.

An equation containing physical quantities expresses a relation which is not generally a mere relation between numbers but be-

tween quantities having dimensions. In every such case the dimensional formula of each term of the equation must be the same. Otherwise, even if there were numerical equality between the two sides of the equation when one set of units is used, this equality would disappear if the same quantities were measured in another set of units.

BOOK I

MECHANICS

CHAPTER I

GENERAL NOTIONS ABOUT FORCES

11. Force. — Man possesses a sense, called the motor sense, by which he is conscious of his exertion of *force*. He finds, by common experience, that by the exertion of force he may move a body, either altogether out of its place, or so as to distort it, or he may simply keep it at rest in some position from which it would move if he were to cease exerting force. From experiences of this last sort especially he has inferred that the body which he keeps at rest is exerting on him a force or *reaction* equal and oppositely directed to the force which he is exerting, and is at rest because of a balance between the forces. He thus infers that bodies exert forces on each other, and when he perceives results taking place without his agency, which he could have produced by exerting force, he ascribes them to forces exerted by other bodies.

When we hold a mass of iron in the hand we are conscious of the exertion of force; and similarly we are conscious of the exertion of force when we pull out a spiral spring or coil of steel wire and keep it stretched. If we suspend the spring by one end and hang the mass of iron on the other end, the spring will be stretched and the mass of iron will come to rest in a position in which it is sustained by the stretched spring. The inference is obvious that the mass exerts a force downward as the hand does when it stretches the spring, and that the spring exerts a force upward as the hand does when it sustains the weight.

12. Scale of Forces. — Our motor sense enables us to discern that forces are not all of equal magnitude and to distinguish which of two forces is the greater when the difference between them is considerable; but it does not furnish any measure of force by which the comparative magnitude of two forces can be expressed. To obtain a quantitative measure of force we need a standard or unit of force and also a scale of forces which are independent of our sense perceptions.

The standard which we shall provisionally adopt is the weight of a certain chosen body or mass of matter. The *weight* of a body is the force which it exerts downward on any body which sustains it. Continued experience has shown that the weight of a permanent body is invariable in time and for any ordinary changes of place.

The kilogram, which has been defined in § 5, serves primarily as a standard mass, and secondarily as a basis for the standard force. We shall adopt the weight of the kilogram as the unit of force. In most English-speaking countries the weight of a pound (§ 5) serves as a unit of force.

By constructing multiples or submultiples of the standard mass any desired multiples or submultiples of the unit force may be obtained, and thus a scale of forces may be constructed.

Copies of the standard kilogram or pound and of their multiples and submultiples are constructed in large numbers, and widely distributed in sets of weights. They are sufficiently well made to enable us to use their weights as the measure of forces in any of the ordinary practical applications of mechanics.

Any force is equal in magnitude to the weight which it will counteract.

13. **Directions of Forces.** — We are conscious that any force which we exert is exerted in some definite direction, and our experience indicates that the same is true of every force. Forces which are equal in magnitude may differ in the effects which they produce and hence be different forces because the directions in which they act are different.

We may indicate the direction in which a force acts by a line drawn in that direction; and if we measure off on that line on any conventional scale a length equal to the magnitude of the force, the limited line thus determined, with the direction in which it is drawn indicated by an arrowhead or in some other way, serves as a pictorial representation of the force.

When a weight is suspended by a fine wire, the upward force which the wire exerts acts along the line of it; and, by conceiving the cross section of the wire to diminish indefinitely while it still sustains the weight, we have it suggested that, as a limit to real conditions and as a useful ideal arrangement, we may conceive of the force exerted by the wire as applied at a point of the body and acting in a line.

The conception which is obtained from this example may be applied to all cases in which forces act. When, as in the case of the weight of a body, the forces seem to be applied continuously

throughout the body, we treat them as if they were applied at points in the body; and further we can find a rule by which any such set of distributed forces can be reduced to a single force acting along a line or applied at a point on it.

We call the point to which we conceive the force to be applied *the point of application*, and the line drawn through the point of application in the direction in which the force acts *the line of action* or *the line of the force.*

14. **Transmissibility of Force.** — Another fact of general observation is of great service in the study of mechanics. It is called the *transmissibility of force.*

When a rope is stretched between two points, the pull applied at one end is transmitted by the rope to the other end, and is there counteracted by an equal and opposite pull. This result occurs because the rope is stretched, and internal forces arise between its parts, which counteract each other along its whole length, and which at the ends counteract the applied forces. But by conceiving the rope to be entirely unyielding, or rigid, we ignore these internal forces, which only arise because of the action of the applied forces, and simply consider either the one or the other of the applied forces as transmitted from one end of the rope to the other, so that the conditions that maintain the rope at rest depend on the two forces only. This being so, it is plain that if a force is given, applied at any point, in a body that may be conceived of as rigid, its point of application may be transferred to any other point in the body, on the line of the force, without thereby altering its effect on the body.

More generally, when three or more forces are applied to an extended body so as to keep it at rest, we may conceive of the body as rigid, and consider that the forces are transmitted through the body in such a way as to produce, at the point of application of each force, an equal and opposite force which counteracts it.

15. **Mechanics.** — In our common experience the motions of bodies and the positions occupied by bodies when they are at rest are due to the actions of forces. The science of *Mechanics* deals with these phenomena in relation with the forces which produce them. Mechanics is usually divided into Statics and Dynamics.

The science of *Statics* deals with the conditions of bodies at rest under the action of forces.

The science of *Dynamics* deals with the motions of bodies produced by forces.

CHAPTER II

STATICS

16. The Problem of Statics. — When a body is at rest under the action of forces, it is said to be in *equilibrium*. Since the body is of importance only as the means of transmitting the forces, it is customary to speak of the forces which act on a body at rest as being themselves in equilibrium. The problem of Statics is to determine the conditions of equilibrium of bodies under the actions of forces.

Principle of Moments

17. The Lever. — The first case of equilibrium which we shall consider is that offered by the *lever*. In one of its forms a lever is a rod or bar which is placed across a fixed support, called its *fulcrum*. When a force is applied to one end of it to turn it about the fulcrum as an axis, the other end applies a force to any body against which it presses. The force at one end is generally in the control of the man using the lever. It is called the *power*. The force applied to the other end is called the *weight*. The ratio of the weight to the power is the *mechanical advantage* of the lever.

It is not convenient to restrict ourselves to the consideration of the lever defined in this way, or to study the determination of the mechanical advantage directly. It is better to generalize the problem at once by considering a body which is fixed by an axis so that its only possible motion is one of turning about the axis, and to which forces are applied at various points and possibly in various directions so as to maintain it in equilibrium. We may then state our problem as being that of determining the conditions of equilibrium of a body free to turn about an axis. The problem of the ordinary lever is plainly a special case of this more general one.

This problem was successfully studied by the Greek geometer Archimedes (287–212 B.C.) and his solution was applied by him to the determination of the center of gravity. His method of discussion is open to criticism for assuming tacitly the very principle which he desires to prove, and we shall not follow it here.

After giving a preliminary definition, which will enable us to state the results of observation more concisely than we otherwise could do, we shall proceed to determine the conditions of equilibrium by a direct appeal to experiment.

18. Moment of Force. — The *moment* of a force about an axis perpendicular to the plane in which the line of the force lies is the product of the magnitude of the force and of the distance of its line from the axis.

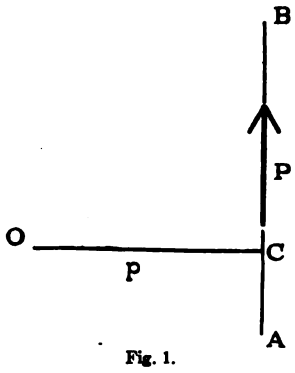


Fig. 1.

Thus, Fig. 1, if we represent by the point O the trace of the axis standing perpendicular to the plane of the paper, by the line AB the line of the force, by the line $OC = p$ the distance of the point O from the line AB, and by P the magnitude of the force, then the moment of the force P about the axis through O is Pp.

According as the force is directed as indicated by the arrow, or is oppositely directed, it tends to turn the body to which it is applied in the counterclockwise sense or in the clockwise sense about the axis. This is often denoted by giving opposite signs to the two moments, using the plus sign with the moment which tends to turn the body in the counterclockwise sense.

19. Equilibrium about an Axis. — In our study of equilibrium about an axis we shall consider only the case in which the axis is perpendicular to the planes in which the forces act; so that each force has a moment in the sense already defined.

To determine the conditions of equilibrium of a body free to turn about an axis, we apply forces to such a body in planes perpendicular to the axis and alter their magnitudes and directions or change their points of application until the body is in equilibrium. We then investigate the relations among the forces in order to detect, if possible, some simple and general condition which holds in every case of equilibrium.

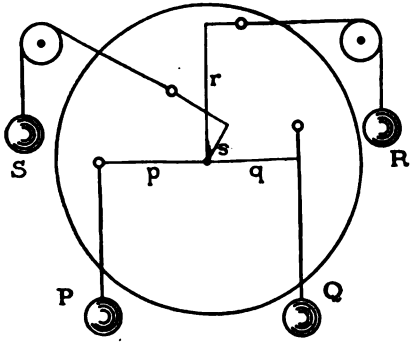


Fig. 2.

A wheel turning easily on a horizontal axle is a suitable body to use in this fundamental experiment. Forces may be applied to it by attaching weights to it at various points by fine cords. The forces act along the cords; when the cords are all vertical, the forces are

parallel; but they may be made to act in any direction by leading the cords aside over smooth pulleys, as indicated in Fig. 2.

It is evident that equal forces arising from other causes may be substituted for the weights without altering the conditions of equilibrium, and that the experiment illustrates the general case of the equilibrium of forces about an axis.

When any case of equilibrium is examined it is found that equilibrium may be destroyed either by changing the magnitude of one of the forces, or by altering its line of action, so that we conclude that the importance of a force in maintaining equilibrium about an axis depends not only upon its magnitude but also upon the position of its line relative to the axis; and proceeding further we find that in every case of equilibrium a simple relation exists among the moments of the forces, namely, that the sum of the moments about the axis of those forces which tend to turn the body in one sense is equal to the sum of the moments of those which tend to turn it in the opposite sense. In the case represented in Fig. 2 we should find

$$Pp + Ss = Qq + Rr.$$

Remembering that it has been agreed in § 18 to give opposite signs to moments which turn the body in opposite senses, we may express the results of our observations in the following statement:

When a body is in equilibrium about an axis, the algebraic sum of the moments about the axis of the forces which act on the body is equal to zero.

In the special case of Fig. 2 we should find

$$Pp - Qq - Rr + Ss = 0.$$

It is convenient at this point to introduce a symbolic method of representing a sum, which is of considerable use in enabling us to abbreviate formulas containing sums of similar quantities. Suppose that we have to add a number of terms which are similar in form but may differ in magnitude and sign. Ordinarily this operation would be represented by writing out the terms, each with its proper sign, so as to form an expression like

$$A + B - C - D + \cdots N.$$

We select one of these terms, say A, to serve as a type of them all, and to this typical term we prefix the letter Σ. This letter indicates the operation of forming all the terms like A, each with its proper sign, and of then adding them together. ΣA therefore expresses the algebraic sum of all the terms. The symbol Σ is generally called the summation sign.

In this notation the law of equilibrium about an axis is expressed by the simple formula

$$\Sigma Pp = 0. \qquad (1)$$

20. The Lever. — Any body of the sort which has been described, which is free to rotate about an axis and which is held in equilibrium by forces, may be called a lever, but the name, as commonly applied, denotes a straight or a bent bar, which is placed on a fulcrum, or arranged so as to be free to turn about an axis perpendicular to its plane, and which is held in equilibrium by two forces. One of these forces is so commonly the weight of a body that it is called the *weight;* the other is the *power*. The distance from the fulcrum or axis to the line of the weight is called the *weight arm;* the corresponding distance to the line of the power is called the *power arm*. Typical levers are represented in Fig. 3.

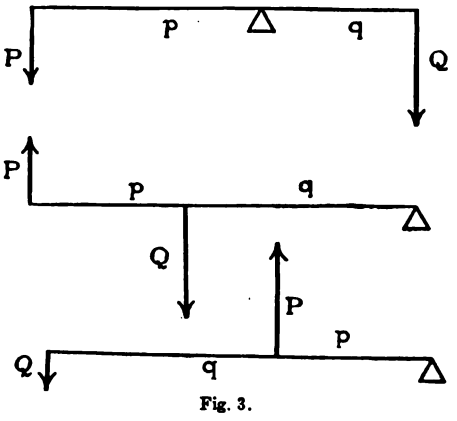


Fig. 3.

Designating the power by P, the weight by Q, the power arm by p, the weight arm by q, the general condition of equilibrium is given by the formula $Pp = Qq$.

The mechanical advantage of the lever is

$$\frac{Q}{P} = \frac{p}{q} \tag{2}$$

and is equal to the ratio of the power arm to the weight arm.

On account of the general use made of it in mechanical operations and in the construction of machines, the lever is known as one of the fundamental *mechanical powers*.

21. Equilibrant and Resultant. — In trying the experiment described in § 19 it is in most cases evident that the axis which supports the body is acted on by a force, and that the body would start off in one direction or another if the axis were removed. We express the fact that the axis prevents the movement of the body, in harmony with our general experience that the action of one force may be neutralized by the opposite action of an equal force, by saying that the axis exerts a counteracting force or *reaction* equal and oppositely directed to the force which the body exerts upon it. The force which the body exerts upon the axis, which results in some way from the combined action of the forces applied

to the body, is called the *resultant* of those forces. We may substitute for the reaction of the axis another applied force, equal and opposite to the resultant, without changing the conditions of equilibrium. The axis may then be removed and the body will be in equilibrium under the action of a system of applied forces. The single force which makes equilibrium with the other applied forces is called the *equilibrant* of the system.

We must again have recourse to experiment to find the relation between the system of applied forces and their equilibrant.

22. Three Parallel Forces in Equilibrium. — This relation is extremely simple in the case of parallel forces, and can be sufficiently

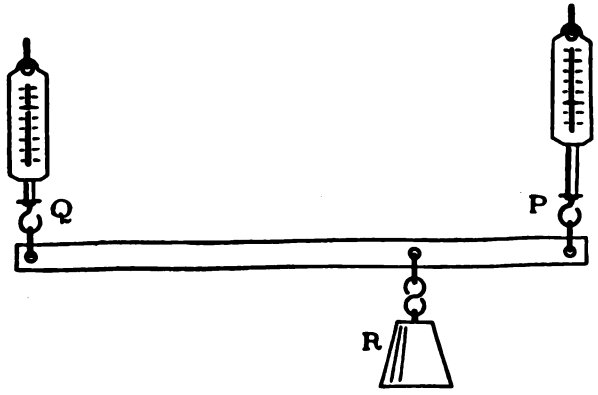


Fig. 4.

established on an experimental basis by the examination of the case in which three such forces are in equilibrium.

The two ends of a light rod, whose weight is either neglected or allowed for, may be suspended by vertical cords from two spring balances so that the rod is horizontal, and a convenient weight may be hung by a cord from various points of the rod (Fig. 4). The two forces exerted upward by the springs and the force exerted downward by the weight form a system of three forces in equilibrium (Fig. 5). Whatever may be the position of the point B from which the weight R hangs, it will always be found that the sum of the forces P and Q is equal to R; and it will be found further that the moments of the forces P and Q about the point B are equal to each other.

In order to state the law most simply it is convenient to consider as positive a force which acts in one direction and as negative a force which acts in the opposite direction. With this convention we may express the law of equilibrium by saying:

When three parallel forces are in equilibrium, they are in one plane; their algebraic sum is equal to zero; and the algebraic sum

of the moments of any two of them about the point of application of the third is equal to zero.

The last of the statements here made is more general than the previous description of the experiment will warrant. That it must be correct may be seen at once if we consider that an axis may be passed through any one of the points A, B, or C without destroying equilibrium, and that then two forces are in equilibrium about that axis, so that, by our first fundamental experiment (§ 19), the sum of their moments must equal zero. We may deduce this general statement from the particular experimental result described. For from that result we have

$$\left.\begin{aligned} P + Q - R &= 0, \\ Pp - Qq &= 0. \end{aligned}\right\} \tag{3}$$

To find the relation which holds between P and R when an axis is passed through the point C, so that Q does not appear as a measured force, we must eliminate Q from these equations. On so doing, writing r for $p + q$, we obtain

$$Pr - Rq = 0, \tag{4}$$

which expresses the equality of the moments of P and R about the point C.

Similarly by eliminating P, we obtain

$$Rp - Qr = 0, \tag{5}$$

which expresses the equality of the moments of Q and R about the point A.

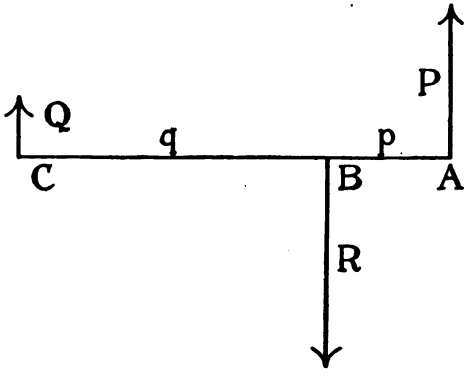


Fig. 5.

Equations 3, 4, 5 express the law of equilibrium of three parallel forces.

Plainly any one of the three forces P, Q, R may be considered the equilibrant of the other two. If the magnitudes and positions of two parallel forces are given, the magnitude and position of the equilibrant may be found by the help of the law of equilibrium, or by the use of the preceding equations.

23. Moment of Three Parallel Forces in Equilibrium about any Axis Perpendicular to their Plane. — If three parallel forces are in equilibrium, the sum of their moments about any axis perpendicular to their plane is equal to zero. This may be inferred directly, for if the three forces are applied to a body, and an axis passed through the body perpendicular to the plane of the forces, the introduction of the axis will not affect the equilibrium of the body, and from the first experimental condition of equilibrium

(§ 19) the sum of the moments of the forces about the axis must equal zero. It may also be proved more formally as follows:

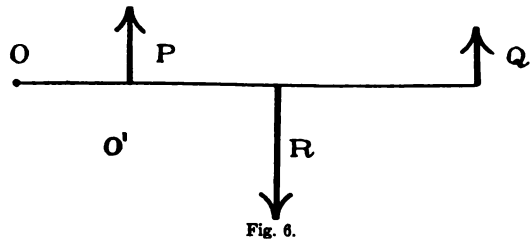


Fig. 6.

Let O (Fig. 6) represent the trace of any axis drawn perpendicular to the plane of the three forces P, Q, R. Let s represent the distance from O to the line of the force R, and p and q, as before, the distances of the lines of the forces P and Q from the line of the force R.

The sum of the moments about the axis through O is

$$P(s-p)+Q(s+q)-Rs,$$

which may be written

$$(P+Q-R)s-(Pp-Qq),$$

and this sum is equal to zero, because from the conditions of equilibrium of three parallel forces we have

$$P+Q-R=0, \quad Pp-Qq=0.$$

The condition expressed by the formula

$$P(s-p)+Q(s+q)-Rs=0 \tag{6}$$

has thus been shown to be a consequence of the equilibrium of the forces. It is therefore a necessary condition of equilibrium. That it is also a sufficient condition may be shown as follows:

Let P, Q, and R be any three parallel forces in a plane, of which we know that the sum of their moments about any axis perpendicular to their plane is equal to zero. Using a, b, c to represent the distances of the lines of the forces P, Q, R (Fig. 6) from the axis through O, and taking the force R as the one whose moment is negative, we write the known condition in the formula

$$Pa+Qb-Rc=0.$$

If a', b', c' represent similarly the distances of the lines of the forces P, Q, R from a parallel axis through any other point O', we have also

$$Pa'+Qb'-Rc'=0,$$

and by subtraction, setting $x=a-a'=b-b'=c-c'$, we have

$$(P+Q-R)x=0,$$

and therefore

$$P+Q-R=0.$$

Also setting $c-a=p$, $b-c=q$, and substituting in the first equation, we have, by the help of the result just obtained, $Qq-Pp=0$. These formulas embody the conditions of equilibrium already established. The condition that the sum of the moments of three forces about any axis perpendicular to their plane shall be equal to zero leads to the conditions of equilibrium established by experiment, and is therefore not only a necessary, but the sufficient, condition of equilibrium.

The condition of equilibrium established in the last paragraph is the simplest example of a general principle of equilibrium known as the principle of moments.

Since the resultant of two parallel forces is simply the equilibrant reversed, it is obtained by finding the force whose line is in the plane of the given forces and whose moment about any axis perpendicular to their plane is equal to the sum of the moments of the given forces.

24. Moment of Three Parallel Forces in Equilibrium about any Axis in a Plane Perpendicular to their Plane. — If three parallel forces are in equilibrium the sum of their moments about any axis in a plane perpendicular to their plane is equal to zero. This proposition may be inferred directly, and proved more formally as follows:

Let the points P, Q, R (Fig. 7) represent the traces of the lines of the forces P, Q, R in equilibrium, and perpendicular to the plane of the paper. Let the line MN represent any axis drawn in the plane of the paper and so in a plane perpendicular to the plane of the forces. This axis will cut the line drawn in the plane of the paper through the lines of the forces, at the point O. Let a, b, c represent the distances of the lines of the forces P, Q, R from the axis MN. Then $Pa + Qb - Rc$ is the sum of the moments of the three forces about the axis MN. Now from similar triangles, if a', b', c' represent the distances of the lines of the forces P, Q, R from the point O, and k a factor of proportion, we have $a = ka'$, $b = kb'$, $c = kc'$, and the sum of the moments becomes

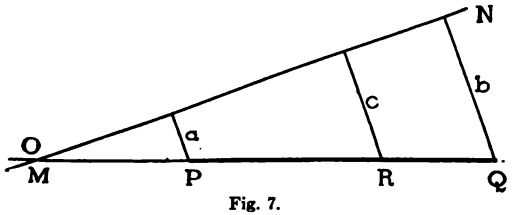


Fig. 7.

$$k\,(Pa' + Qb' - Rc').$$

But by the previous proposition (§ 23) the sum in parenthesis is equal to zero if the forces are in equilibrium, so that the sum of the moments about the axis MN is also equal to zero.

By an argument essentially similar to that used in § 23 it may be shown that the condition that the sum of the moments of three forces in a plane about any axis perpendicular to their plane is equal to zero is not only a necessary, but the sufficient, condition of equilibrium of the forces. This proposition is a more general example of the application of the principle of moments.

25. Equilibrium of Many Parallel Forces. — By the rules now obtained we may find the resultant of any two given parallel forces. This resultant is in all respects, as regards maintaining equilibrium, equivalent to the given forces, and so may be combined with a

third parallel force, and a resultant found which is the resultant of the three given forces. If more forces are given, this process may be continued step by step until all the forces have been combined and their final resultant determined. This will be a force of definite magnitude and direction and will have a definite position among the given forces. Reversed, it is the equilibrant of the system of forces.

This is not ordinarily the most convenient method of finding the equilibrant. To develop the usual method we consider first the case of two parallel forces. If we are given simply two forces, P and Q, and their distances x_p and x_q from an axis, the rule stated in § 24 is sufficient to find the distance x_r of their equilibrant R from the axis, but does not fix the position of the line of R. We may get this position if we know the positions of the lines of P and Q, for since we know that the line of the equilibrant cuts the straight line joining the lines of P and Q, we can locate the line of the equilibrant on that point of this line which is at the distance x_r from the axis. Ordinarily we are given the coördinates of the lines of P and Q referred to rectangular axes of x and y (Fig. 8). We then can form three equations which contain the complete solution of the problem:

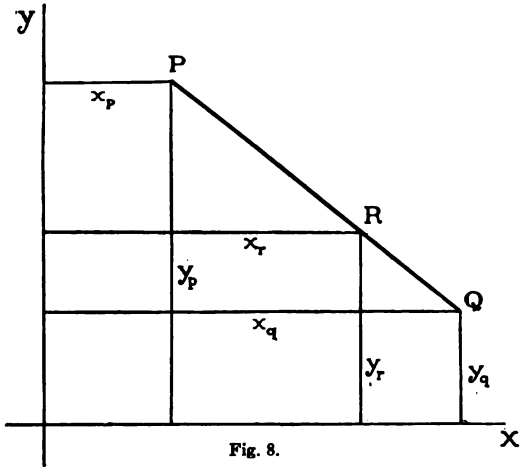


Fig. 8.

$$\left.\begin{aligned} P + Q - R &= 0, \\ Px_p + Qx_q - Rx_r &= 0, \\ Py_p + Qy_q - Ry_r &= 0. \end{aligned}\right\} \tag{7}$$

These equations give us the value of R and the coördinates x_r and y_r of its line of action.

If many parallel forces are given in a similar way, we can proceed step by step to find the final equilibrant, as in the method described in the first paragraph. But in this case the process may be much simplified. To illustrate this, let us consider that another

force S is given, besides the forces P and Q. The force R is the resultant of P and Q. The equilibrant R' of R and S is found from the equations

$$R + S - R' = 0,$$
$$Rx_r + Sx_s - R'x_{r'} = 0,$$
$$Ry_r + Sy_s - R'y_{r'} = 0.$$

But from the equations defining R we may substitute the values of R, Rx_r, Ry_r, and obtain

$$P + Q + S - R' = 0,$$
$$Px_p + Qx_q + Sx_s - R'x_{r'} = 0,$$
$$Py_p + Qy_q + Sy_s - R'y_{r'} = 0.$$

These equations are sufficient to determine the final equilibrant R' and its coördinates $x_{r'}$ and $y_{r'}$.

Their law of formation is evident, and will hold no matter how many forces are given or in what order they are taken. We may therefore write the equations for the equilibrant R of any number of parallel forces, using the symbolic notation described in § 19, in the forms

$$\left.\begin{aligned} \Sigma P - R &= 0, \\ \Sigma Px_p - Rx_r &= 0, \\ \Sigma Py_p - Ry_r &= 0. \end{aligned}\right\} \tag{8}$$

The equations of equilibrium of any system of parallel forces in equilibrium are simply these equations with the equilibrant set equal to zero. They therefore become

$$\Sigma P = 0;\ \Sigma Px_p = 0;\ \Sigma Py_p = 0. \tag{9}$$

26. The Couple. — A simple modification of Equation 5, which gives the position of the equilibrant of two oppositely directed forces, enables us to consider the interesting special case in which the two forces are equal. If we represent by d the distance between the lines of the forces R and Q (Fig. 9) we may write Equation 5 in the form

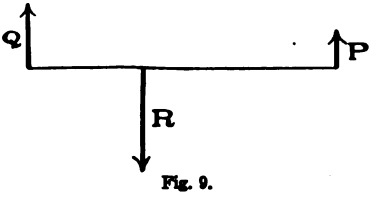


Fig. 9.

$$Rp - Q(p + d) = 0,$$

and obtain the distance p of the equilibrant P from the line of R from

$$p = \frac{Qd}{R - Q}.$$

The magnitude of the equilibrant P is the algebraic sum of the two forces, or

$$P = R - Q,$$

and it is oppositely directed to the greater force.

The special case in which $R = Q$ is the limit of this case as R approaches Q in value.

In the limit we have

$$p = \infty, \quad P = 0;$$

that is, the conditions of equilibrium require an equilibrant equal to zero applied at a point infinitely distant from the forces. We conclude that the pair of forces considered cannot be equilibrated by any single force.

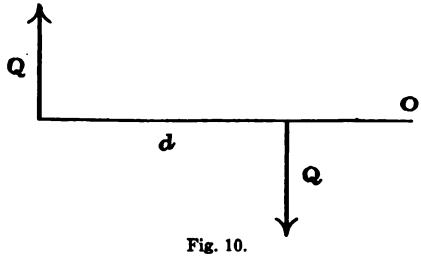


Fig. 10.

Such a pair of forces, equal and oppositely directed, and not in the same line, is called a *couple*. The sum of the moments of the two forces, about any point O in their plane (Fig. 10), is the constant quantity Qd, the product of either of the forces and the perpendicular distance between the lines of the forces. This product is called the *moment of the couple*. The distance between the lines of the forces is called the *arm of the couple*.

Since a couple has no equilibrant, it can only be equilibrated by an oppositely acting couple of equal moment. Since the moment of the couple is independent of the axis of rotation, all couples of equal moment are equivalent in their effect on a body, wherever in the body they may be applied.

27. The Pulley. — The simplest illustration of the equilibrium of parallel forces is afforded by another of the mechanical powers, the *pulley*.

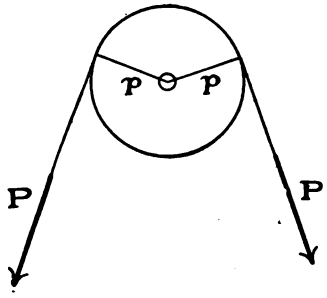


Fig. 11.

The pulley is a wheel turning freely on an axle. Over a groove in the rim of the wheel is passed a flexible cord. Weights may be hung on the two ends of the cord or forces may be applied to them. If the supports of the axle are fixed, the pulley is called a *fixed pulley* (Fig. 11). Whether the two sections of the cord are parallel or not, it is plain

from symmetry that the fixed pulley will be in equilibrium when the forces applied to the ends of the cord are equal. This condition of equilibrium may be otherwise readily deduced, by considering the points on the wheel at which the cords leave it as the points of application of the forces which are applied to the cord, and considering the radii of the wheel drawn from the axis to those points as two lever arms of equal length. The moments of force on either side of the axis are then plainly equal when the weights or forces are equal.

The mechanical advantage of the fixed pulley is 1, and its use is simply to change the direction of the force which is applied to the cord.

Sometimes the pulley is not fixed on a support but is hung (Fig. 12) in a loop of the cord, one end of which is fastened to a fixed

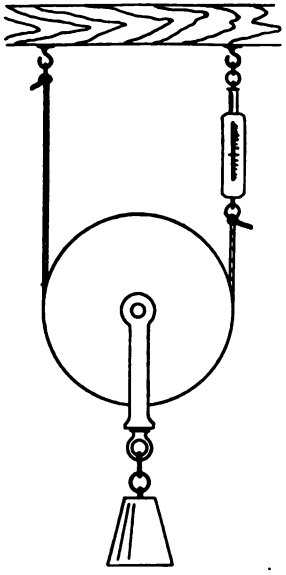


Fig. 12.

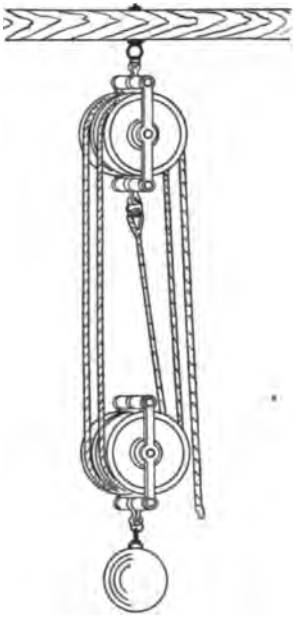


Fig. 13.

support, while a force is applied to the other end. A force is applied to the framework in which the axle of the pulley is supported. Such a pulley is called a *movable pulley*. The conditions of equilibrium may be deduced by treating the pulley as a body to which three parallel forces are applied. Two of these forces are

the tensions in the cords at the points where they leave the rim of the pulley; and if the pulley is smooth, as we shall assume, they are equal. The third force, or equilibrant, of these is the force applied in the opposite direction to the axle of the pulley; and this is equal to the sum of the tensions, or is double of either of them or of the force applied to the free end of the cord which causes the tension. The mechanical advantage of the movable pulley is therefore 2.

Combinations of several pulleys are often employed to obtain a larger mechanical advantage. In the *block and tackle*, which is a typical one of these combinations, several pulleys are mounted in a framework, or block, on the same axle (Fig. 13). One of these blocks is attached to a fixed support, and a cord attached to the block is carried over one of the pulleys of a similar free or movable combination and back over a pulley of the fixed combination and so on until the cord has passed over all the pulleys. A force is applied to the free end and makes equilibrium with a force applied to the axle of the movable block. The branches of the cord are generally not strictly parallel, but a first approximation to the conditions of equilibrium may be made by treating them as if they were parallel. The applied force P and the equal reaction at the point of attachment R set up a tension equal to P in all parts of the cord. If there are n pulleys in the block the tension in the cord applies to the block $2n$ forces each equal to P, and the equilibrant Q is equal to $2nP$. The mechanical advantage of the block and tackle is then approximately $\frac{Q}{P} = 2n$. This is numerically equal to the number of cords that support the movable pulley.

28. Center of Gravity. — A more general application of the results which we have attained will lead to the determination of the *center of gravity*. In fact it was for this purpose that Archimedes investigated the properties of the lever.

The weight of a body is equal to the sum of the weights of its parts, and these weights are parallel forces. When the body is in any position, we may find the line of their resultant by the rule developed in § 25. If the body is turned into another position, a new line of their resultant may be found; and so for each position in which the body is placed. In each new position the weights are still parallel forces and their points of application have the same relative positions in the body; but the relations of these points to the horizontal axes of reference change for each new position; and

simple inspection will not show that the lines of the various resultants found will have any connection with each other. More extended analysis than can conveniently be presented here leads to the conclusion that the lines of the resultants all meet at a common point. This point is evidently one at which we may conceive the weight of the body applied so as to produce the same effect as the actual distributed weights of the parts of the body. We accordingly name it the *center of gravity*, and define the center of gravity of a body as the common point of intersection of the resultant weights of the body in different positions.

If a body is freely suspended by a thread or poised on a point, the line of the equilibrant passes through the center of gravity. If two or more such lines can be traced, they will be found to intersect at the center of gravity. If a body is poised on a point placed at the center of gravity, it will be in equilibrium in any position about that point.

Manifestly any set of bodies will have a center of gravity, which is to be found by the same method as that used for a single body. It is often convenient to find first the centers of gravity of the several bodies of the set. These are then treated as points to which the weights of the bodies are applied, and their common center of gravity is found.

The sum of the moments of the weights of all parts of a body or set of bodies will be equal to zero about any horizontal axis drawn in a vertical plane which contains the center of gravity.

In the actual calculation of the center of gravity of an extended body, we treat the body as if it were made up of a continuous distribution of matter, and we determine the sum of the moments and the sum of the weights by the use of the integral calculus. Special cases can often be solved from symmetry or by special devices.

The center of gravity of symmetrical or regular figures, such as a limited straight line, a rectangular parallelepiped, a right cylinder, or an ellipsoid, lies at the center of figure.

29. Stable and Unstable Equilibrium. — When a body, like a box or a table, is at rest on a level surface, it will be found that, if a perpendicular is let fall from its center of gravity to the surface, the foot of the perpendicular will lie inside the closed figure formed by drawing lines through the points on which the body is supported. In this case it will be seen at once that if the body is tilted about any line passing through two of its points of support, the weight of the body introduces a moment of force about that line, which opposes the motion. Such a motion, therefore, is con-

trary to that which the weight of the body would occasion, and will not arise from the action of the weight. The body is said to be in *stable equilibrium.*

From another point of view, which will be considered more fully later, any possible motion of the body, such as would tilt it over, involves lifting the center of gravity. We may conclude from general experience that the center of gravity of a body never spontaneously rises, and that in the circumstances the body will not move; and further, that even if slightly tilted, it will return to its original position.

In some cases the perpendicular let fall from the center of gravity meets the level surface on which the body rests at a point in the line joining the points of support, or at a point of support. These cases are those of unstable or neutral equilibrium. In *unstable equilibrium*, like that of a block balanced on an edge, or of a cone balanced on its apex, any tilting of the body about the line or point of support introduces a moment of force due to the weight which is in the same sense as the motion and tends to increase it. From the other point of view, the motion which tilts the body lowers the center of gravity, so that the body, if once tilted ever so slightly, will not return to its original position but will recede further from it.

In *neutral equilibrium*, like that of a cylinder lying on its side, or of a sphere, any tilting of the body changes the position of the line or point of support, so that it is always at the foot of the perpendicular let fall from the center of gravity. No turning moment is introduced, and the center of gravity neither rises nor falls. The body remains at rest in its new position.

30. **Principle of Moments.** — Before leaving the subject of moments, we shall give a more general statement of the criterion of equilibrium known as the *principle of moments.* If any number of forces, whether parallel or not, are in equilibrium, the sum of their moments about any axis is equal to zero. This we can assert on the basis of our fundamental experiments, at least for forces which lie in a plane or in parallel planes. Conversely, if the sum of the moments of a set of forces about any arbitrarily chosen axis is equal to zero, the forces are in equilibrium. This condition, in fact, is a sufficient as well as a necessary condition of equilibrium. By a method similar to that used in § 24 it can be shown to lead to the particular conditions of equilibrium which are embodied in Equations 9. The general condition of equilibrium here stated is the principle of moments.

EXAMPLES, I

1. We are given a bent lever (Fig. 14), one branch of which, of length p, is horizontal; the other branch, of length q, makes an angle α with the horizontal; the fulcrum is at the angle; a weight P is hung on the horizontal branch, and balanced by a weight Q on the end of the other branch; find the formula for Q. The moment of Q about the fulcrum is $Qq\cos\alpha$, and this is equal to Pp when equilibrium obtains. The length $q\cos\alpha$, which replaces in a straight lever the actual length q by which equilibrium is made, was called by Leonardo da Vinci (1452–1519) the potential lever.

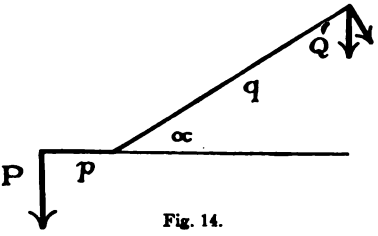


Fig. 14.

2. Two boards each of length d lie end to end on a base, and are hinged at their outer ends. The two ends of a rope are attached to the inner ends, and

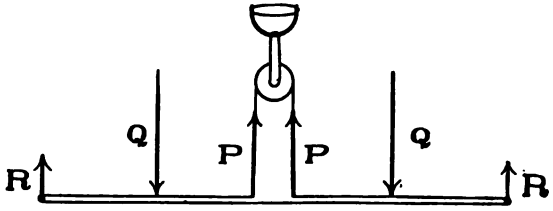


Fig. 15.

the rope passed over a pulley. A man of weight W standing with his feet placed on the boards, at equal distances from their ends, lifts up on the pulley. Find the force that he must exert to just raise the free ends of the boards (Fig. 15).

The forces to be considered are the same for each board. They are P, the upward pull, applied at the free end; R, the unknown reaction at the hinge, and Q, the force exerted against the board by the foot. If the man had no weight, his pull on the rope would be equal to the push of his feet against the board; therefore the force Q is given by $Q = P + \frac{W}{2}$. If we consider the board as a lever, with the fulcrum at the hinge, and set up the equation of moments about that fulcrum, we have, using l to represent the distance of the line of Q from the fulcrum, $Ql - Pd = 0$, and substituting for Q, we get

$$P = \frac{Wl}{2(d-l)}.$$

The reaction R is given by $R + P - Q = 0$, and hence $R = \frac{W}{2}$.

From these formulas we find the changes in P as the feet change positions on the boards.

3. A man carries a light pole, whose weight may be neglected, horizontally over his shoulder. A weight P is hung to the end of the pole, and balanced by the hand applied at the other end. The length of the pole is l, and the distance

of the hand from the shoulder is d. Find the pull exerted by the hand and the reaction of the shoulder.

The forces are Q and R exerted by the hand and by the shoulder. The equations of equilibrium are $P + Q - R = 0$; $Qd - P(l - d) = 0$. We get

$$Q = \frac{P(l-d)}{d}, \quad R = \frac{Pl}{d}.$$

4. A bridge of uniform structure, having a weight W, rests on two piers at its ends, and a weight P is placed on it; find the reactions at the piers. The length of the bridge is l, the distance of the weight from one end is d.

Taking moments in turn about each point of support, the equations of equilibrium are $R_1 + R_2 = W + P$, $R_2 l = \frac{Wl}{2} + Pd$, $R_1 l = \frac{Wl}{2} + P(l - d)$.

We get $R_1 = \frac{W}{2} + P\frac{l-d}{l}$, $R_2 = \frac{W}{2} + \frac{Pd}{l}$.

5. Two boys, one of weight P, the other of weight Q, are setting up a seesaw with a plank l feet long and of weight W. Where must the support be placed so that the plank will be balanced with the boys on its ends?

The resultant $R = Q + P + W$ must pass through the fulcrum. Let x represent its distance from the end at which the weight Q is placed. Then $Rx = \frac{Wl}{2} + Pl$.

6. A balance beam is generally mounted on a knife-edge so that, when there are no weights in the scale pans, the center of gravity of the beam is a little below the point of support. Find the equation giving the sensibility of the balance, that is, the deflection of the beam from the horizontal when the weights in the pans differ by a small known amount.

We assume that, as would be the case in an ideal adjustment, the knife-edges sustaining the scale pans and the knife-edge sustaining the beam are in the same line. In this case, if we call the half length of the beam l, and the distance of its center of gravity below the point of support d, the weight of the beam W, the weights in the scale pans P and $P + p$, where p is a small additional weight; and if we suppose the beam to incline from the horizontal by a small angle β (Fig. 16), we may write the equation of equilibrium $Wd \sin\beta + Pl \cos\beta = (P + p) l \cos\beta$, or

$$\tan\beta = \frac{pl}{Wd}.$$

The sensibility, which is measured by $\tan\beta$ for a given value of p, is then proportional to the length of the beam, and inversely proportional to the distance of the center of gravity below the knife-edge.

If the points at which the weights are applied to the beam and the sustaining knife-edge are not in one line, the sensibility depends also upon the weight P.

7. A table with a heavy circular top stands on three similar heavy legs, placed at the angles of an equilateral triangle. The edge of the top passes through the angles of this triangle. Find the lightest weight that will tilt the table.

The weight must be placed (Fig. 17) at a point like O, on the edge and midway between two legs. Call this weight P, the weight of the top W, the weight of a leg Q, the radius of the top r.

The distance from O to the line AB is $r/2$. The equation of moments about the line AB is $\frac{Pr}{2} = \frac{Wr}{2} + \frac{Q3r}{2}$; hence $P = W + 3Q$.

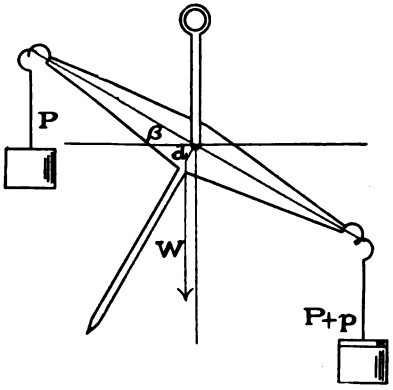


Fig. 16.

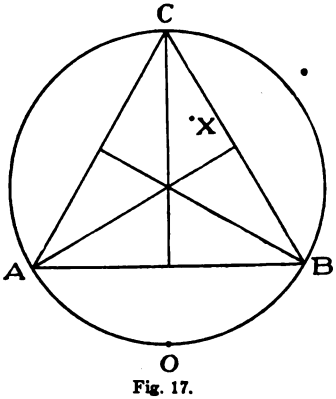


Fig. 17.

8. If a weight P rests on the table of Example 7, at a point X whose position relative to the legs can be determined by measurement, find the reactions on the legs.

Call the distance from X to the line CB, a, the distance from X to the line AC, b. Then we get

$$R_A = \frac{W}{3} + Q + \frac{2Pa}{3r}, \quad R_B = \frac{W}{3} + Q + \frac{2Pb}{3r}, \quad R_C = \frac{W}{3} + Q + P\left(\frac{3r - 2(a+b)}{3r}\right).$$

Calling the distance from X to the line AB, c, we get from the value of R_C, $3r = 2(a + b + c)$, which can be proved otherwise by setting the area of the equilateral triangle equal to the sum of the three triangles which stand on the sides of the equilateral triangle as bases, and whose altitudes are a, b, c.

9. Find the center of gravity of a plane triangle.

A line drawn from an apex to the mid-point of the opposite base bisects all the lines parallel with that base, and therefore contains the center of gravity. The center is at the point where such lines drawn from the three apices of the triangle intersect.

Principle of Work

31. The Inclined Plane. — We shall next consider the conditions of equilibrium of a body on an inclined plane. An *inclined plane* is a perfectly smooth rigid plane, inclined to the horizontal at any angle. A weight may be supported on it by a force directed upward parallel with the surface of the plane (Fig. 18). The inclined plane is one of the mechanical powers. We shall investigate the problem of finding its mechanical advantage, that is, the ratio of the weight on the plane to the force which will sustain it.

It is evident that the inclined plane is a special case of a more general arrangement in which a body is kept in equilibrium on a smooth plane by two forces, one of which is parallel to the plane,

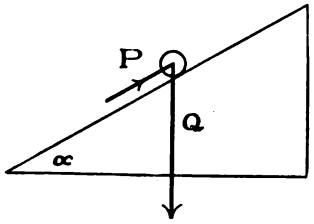


Fig. 18.

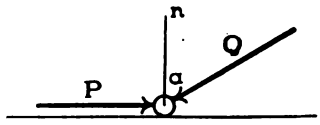


Fig. 19.

while the other makes some angle α, not greater than a right angle, with the normal to the plane (Fig. 19).

An interesting solution of this problem, which rests directly upon fundamental perceptions, and does not involve the use of any formulated principle of mechanics, was given by Stevinus (1605). Stevinus's solution is as follows: Let us suppose two inclined planes such as the one which has been described, starting from the same horizontal plane, and meeting at a summit. Let us suppose, further, a perfectly smooth and flexible, heavy, endless cord hung over the planes (Fig. 20). Every one will admit on inspection of this arrangement that, if the cord is originally at rest, it will remain at rest, that is, its weight will be so distributed that the cord

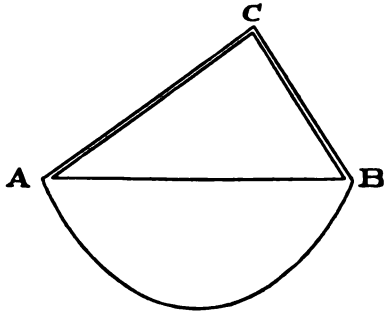


Fig. 20.

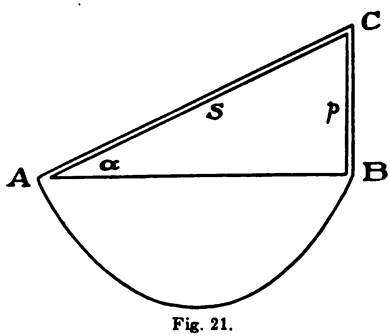


Fig. 21.

will be in equilibrium. If this were not so, the cord would have of itself the power of initiating and maintaining continuous motion around the inclined planes. The part of the cord which hangs below the lowest points of the planes is symmetrical with respect to those points, so that equilibrium must be maintained by those

parts of the cord which rest on the two planes. Since equilibrium is maintained, the forces which these two parts exert in opposite directions are equal. Now the weights of the two parts are proportional to their lengths; and therefore the weights which will be sustained by equal forces on planes of the same height are proportional to the lengths of the planes.

If one of the two planes is vertical (Fig. 21), the force exerted by the part of the cord which hangs vertically is the weight of that part. It will sustain on the inclined plane a part of the cord equal to the length of the inclined plane. The ratio of the weight of the inclined part CA of the cord to the force which sustains it, that is, to the weight of the vertical part CB of the cord, is the mechanical advantage of the inclined plane. The mechanical advantage is therefore equal to the ratio of the length of the plane to its height.

If we represent by p the height of the inclined plane, and by s its length or slant height, we may express the mechanical advantage of the plane, or the ratio of the weight Q lying on the plane to the force P which sustains it, by the formula

$$\frac{Q}{P} = \frac{s}{p}. \tag{10}$$

This formula expresses the condition of equilibrium of a weight on an inclined plane.

Since $p/s = \sin\alpha$, if α represents the angle of inclination of the plane, the condition of equilibrium may also be expressed by the formula

$$P = Q \sin\alpha. \tag{11}$$

The force Q may be any other force than the weight of the body; but the condition of equilibrium will still be expressed by the same formula.

32. Galileo's Solution of the Problem of the Inclined Plane. — Another solution of the problem of the inclined plane was given by Galileo, in which he perceived and applied a fundamental condition of equilibrium which is now known as the principle of work. To exhibit Galileo's solution we suppose that the weight which rests on the inclined plane is sustained by a cord passing over a pulley at the top of the plane, and stretched by a suspended weight suitably adjusted to maintain equilibrium. On contemplating this arrangement Galileo perceived that the only spontaneous motion of a body or system of bodies upon which the forces which act are the weights of its parts, must be one in which the center of gravity

of the system sinks; and he perceived that the condition of equilibrium of such a system must therefore be that the only possible motions are such that they do not lower the center of gravity. In the case of the two bodies in equilibrium on the inclined plane he perceived further that since motion either up or down the plane is equally possible, so that if a motion in one direction raises the center of gravity a motion in the opposite direction will lower it, the condition of equilibrium is that the only possible motions of the system are such that the center of gravity neither rises nor sinks. By accepting this principle as the expression of the condition of

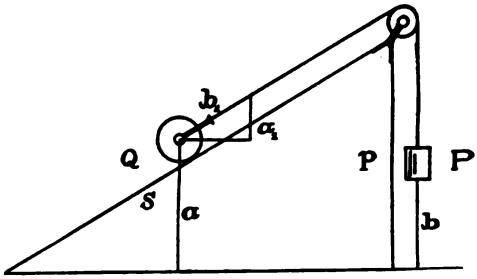


Fig. 22.

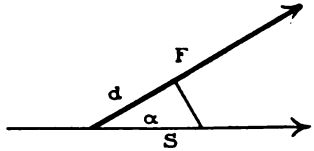


Fig. 23.

equilibrium he arrived at the same relation between the weight and the applied force as that which we have obtained by the method of Stevinus.

We may apply Galileo's principle as follows: Let us represent (Fig. 22) by a and b the heights of the weights Q and P respectively above the horizontal plane. The height c of the center of gravity is then given by $c = \frac{Qa + Pb}{Q + P}$. Now let the weight Q slide up — or down — the plane through the small distance b_1. The weight P will fall — or rise — vertically through the same distance. The vertical distance through which the weight Q rises is a_1, which is related to b_1 by the proportion $a_1 : b_1 = p : s$, where p and s represent the height and length of the plane respectively. The height c_1 of the center of gravity will then become

$$c_1 = \frac{Q(a + a_1) + P(b - b_1)}{Q + P},$$

and if this height is to remain unchanged we must have $c = c_1$, or

$$Qa + Pb = Q(a + a_1) + P(b - b_1)$$

or $$0 = Qa_1 - Pb_1$$

or $$Q : P = b_1 : a_1 = s : p,$$

so that the mechanical advantage is again found to be $\frac{Q}{P} = \frac{s}{p}$.

33. Work. — When the point of application of a force F is moved in the direction of the force through any distance s, the product Fs of the force and the distance through which the point of application moves is called the *work* done by the force. Furthermore, if the motion of the point of application is not in the line of the force but is inclined to it at any angle α (Fig. 23), the work done by the force is measured by $Fs\cos\alpha = Fd$, the product of the force and the projection upon the line of the force of the distance through which the point of application moves. The value of $\cos\alpha$ may range from $+1$ to -1. According as it is positive or negative, the work done by the force is positive or negative.

This definition of work must be taken for the present as an entirely arbitrary one. It is justified by the consistency and success with which it may be applied in all branches of mechanics.

34. Principle of Work. — In working out the problem of the inclined plane by Galileo's method, we obtained the formula $Pb_1 = Qa_1$, and an inspection of the diagram (§ 32) will show that these two products measure respectively the positive work done by the weight P and the negative work done by the weight Q. These two amounts of work are numerically equal, and hence their algebraic sum is equal to zero. This result is a consequence of Galileo's principle and is in fact another and more general form of it; for it obviously is not limited in its application to the special case in which the forces are weights. They may be any two forces acting over a smooth plane, one inclined to the plane and the other parallel to it.

We may state the *principle of work* in the form in which we can use it by saying that a system of forces is in equilibrium when for any small movement of the system which is consistent with the conditions of the system the algebraic sum of the amounts of work done is equal to zero. If we represent by P any one of the forces, by s the displacement of its point of application, and by α the angle between the line of the force and the line of motion of its point of application, we may express the condition of equilibrium of a system by the formula

$$\Sigma Ps\cos\alpha = 0. \qquad (12)$$

This principle was stated in this general form by John Bernoulli in a letter to Varignon (1717) and published in Varignon's treatise on Mechanics in 1725. The same principle was announced by Newton in the *Principia*, but was not used by him in the solution of problems.

35. Application to Case of Lever. — We may test the principle of work by applying it to the lever. To take the simplest case first

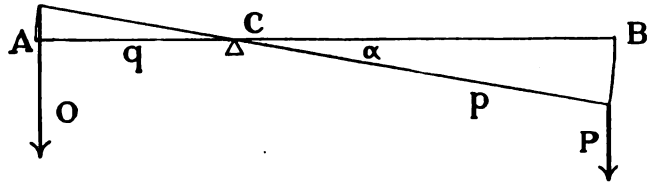


Fig. 24.

let us consider the lever straight and horizontal (Fig. 24). It is in equilibrium under the action of the forces P and Q. Suppose it tilted through the infinitesimal angle α. The end A will be lifted through the distance $q\alpha$, and the end B will be lowered through the distance $p\alpha$. The sum of the two amounts of work done will be

$$Pp\alpha - Qq\alpha,$$

and may be set equal to zero by the principle of work.

Hence we have

$$Pp - Qq = 0,$$

which is the condition of equilibrium expressed in terms of the moments.

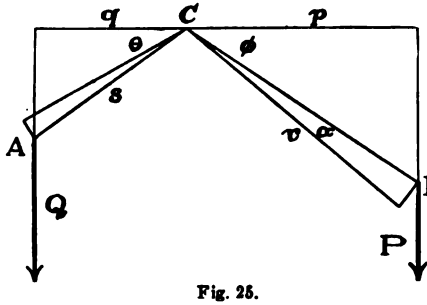


Fig. 25.

In the case of the bent lever (Fig. 25) whose branches or parts make the angles θ and ϕ respectively with the line perpendicular to the lines of the forces, we suppose the lever tilted through the infinitesimal angle α. The distance through which the end A moves is $s\alpha$, that through which the end B moves is $r\alpha$. Since the lines of these motions are perpendicular to the lines s and r respectively, they make the angles $(\pi - \theta)$ and ϕ respectively with the lines of the forces Q and P. Hence the work done by the force Q is $Qs\alpha \cos(\pi - \theta) = -Qs\alpha \cos\theta$, that done by the force P is $Pr\alpha \cos\phi$, and the principle of work gives

$$Pr\alpha \cos\phi - Qs\alpha \cos\theta = 0.$$

From this we get

$$Pr \cos\phi - Qs \cos\theta = 0;$$

or, since

$$r \cos\phi = p,\ s \cos\theta = q,$$

$$Pp - Qq = 0,$$

which is again the condition of equilibrium expressed in terms of the moments.

36. Application to Case of Pulley. — In the case of the block and tackle, with the block containing n pulleys, we see by inspection of the system (Fig. 12) that if a length s of cord is drawn off at the free end, each of the $2n$ branches of the cord between the two

blocks is shortened by an equal share of the whole length, so that the movable block moves toward the fixed one through the distance $\frac{s}{2n}$. Applying the principle of work, we expect equilibrium when

$$Ps - \frac{Qs}{2n} = 0,$$

or when

$$2nP = Q.$$

This is the condition of equilibrium obtained by the use of the principle of moments.

37. Compound Levers. — The principle of work is often capable of furnishing easy solutions of problems which are complicated and difficult to handle by the principle of moments.

As an illustration, let us consider the case of two levers so connected that the weight end of the one is applied to the power end

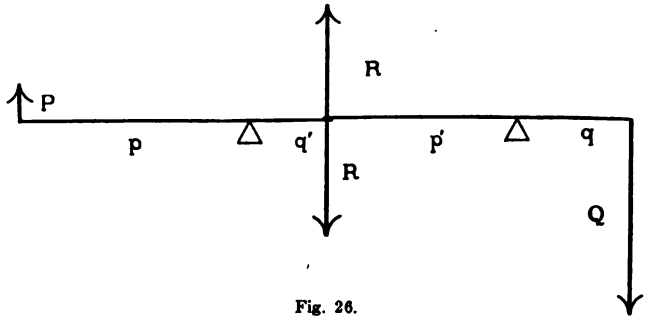


Fig. 26.

of the other (Fig. 26). This pair of levers will be in equilibrium when a force is applied to the power end of the first one and a weight of the proper magnitude to the weight end of the second one. The ratio between this weight and the power can be worked out by the aid of the principle of moments applied successively to the two levers, on the supposition that, when the first lever is in equilibrium, the force which its weight end applies to the power end of the second lever is that force which will maintain the first lever in equilibrium by itself. With the forces and lever arms as indicated in the diagram, the principle of moments applied to each lever gives

$$Pp = Rq', \; Rp' = Qq,$$

and by the elimination of R we obtain, as the condition of equilibrium of the forces P and Q,

$$Ppp' = Qqq'.$$

Equilibrium will exist when the product of the power and the power arms is equal to the product of the weight and the weight arms. In determining the conditions of equilibrium of this system, or of any more complicated system, by the principle of moments, our knowledge of the several levers which make up the system must be complete.

Let us now apply the principle of work to a system of interconnected levers of any degree of complication. When we consider the work done as the system of levers undergoes a small displacement, we perceive that the forces which are exerted between two levers of the system are equal and opposite at the point of junction, so that no work is done when that point moves. The same being true for every point of junction, the only work that is done is that done by the forces applied to the two free ends of the system. Equilibrium will exist when the work done by one of these forces is equal and opposite in sign to that done by the other force. The ratio of the weight to the power is therefore the inverse ratio of the distances through which the weight and the power move respectively, and to determine the conditions of equilibrium, no knowledge of the intermediate structure of the system is necessary. If we know simply the distances through which the two ends of the system move, when the system undergoes a small displacement, we are able to determine the conditions of equilibrium.

EXAMPLES, II

1. *To discuss the problem of the inclined plane by the principle of moments.* Galileo confirmed his confidence in the principle of work, which he used in solving the problem of equilibrium on an inclined plane, by reducing the problem to a form in which it could be handled by the principle of moments.

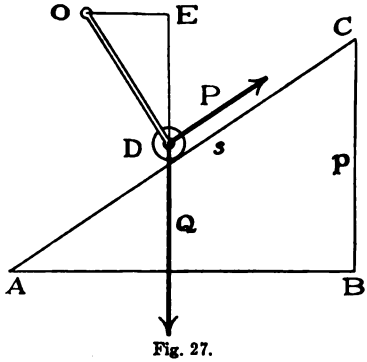


Fig. 27.

His demonstration (Fig. 27) supposes the body on the plane to be attached to an arm *OD* which is pivoted to turn about an axis at *O* and which stands perpendicular to the plane.

The arm permits an infinitesimal motion of the body up or down the plane, and so does not disturb the conditions of equilibrium, which still depend upon the relative magnitudes of the forces *P* and *Q*. Then if these forces keep the body in equilibrium on the plane, they will still maintain equilibrium if the plane is removed and the body is supported by the

arm. By the principle of work we found the forces to be related by the proportion $P:Q = p:s$.

Now the triangle ODE described by drawing OE perpendicular to the line of the weight Q, and DE in the line of the weight, is similar to the triangle ABC, and therefore $OE : OD = p : s$.

Using this relation we obtain $P: Q = OE : OD$, or $P \cdot OD - Q \cdot OE = 0$, which expresses the condition that the moments of the two forces are equal, and that the forces are in equilibrium, about the axis through the point O.

2. *Roberval's Balance.*

A parallelogram of four pieces, jointed at the corners, is supported (Fig. 28) on an upright post by two pins passing through the middle points of the two horizontal pieces. The two end pieces carry pans for the reception of weights. Show that the system will be in equilibrium when equal weights are placed anywhere in the pans.

Any displacement will lift one pan as much as it lowers the other one.

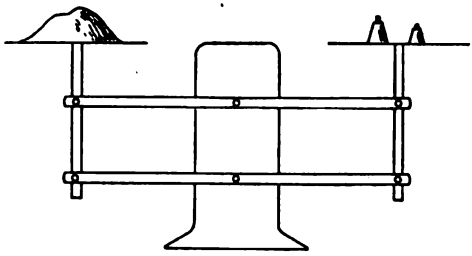


Fig. 28.

3. *A heavy cube rests on a horizontal plane. Find the least force which is needed to tilt it over.*

The weight of the cube is W, the least force P, the side of the cube a. If P is applied to an upper edge perpendicularly to the plane containing it and the opposite lower edge, it will be the least force which will have a moment equal to that of the cube about the lower edge. The equation of moments gives

$$P\sqrt{2}a = W\frac{a}{2}, \text{ or } P = \frac{\sqrt{2}}{4}W.$$

4. *What force P parallel with the base of an inclined plane of inclination ϕ will make equilibrium with a weight W on the plane?*

If the weight moves up the plane through a distance s, the work done by P is $Ps\cos\phi$, the work done by W, $Ws\sin\phi$. In case of equilibrium $Ps\cos\phi = Ws\sin\phi$, whence $P = W\tan\phi$.

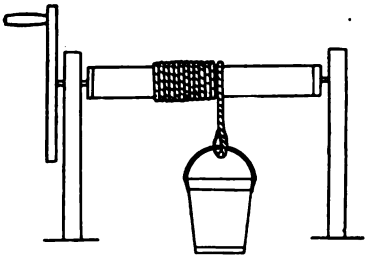


Fig. 29.

5. *Find the mechanical advantage of the wheel and axle.*

The wheel and axle is a general name for that class of mechanical powers in which a rope, by which the weight is lifted or an equivalent force applied, is wound up on an axle or cylinder by turning it with a lever (Fig. 29).

Call the radius of the wheel, or the length of the lever, R, and the radius of the axle r. When the cylinder revolves once, the force P applied to the lever does work $P2\pi R$; the work done on the weight W is $W2\pi r$. Setting these equal we have $\frac{W}{P} = \frac{R}{r}$. The mechanical advantage of the wheel and axle is the ratio of the two radii.

6. *Find the mechanical advantage of the differential wheel and axle.*

The differential wheel and axle is like the wheel and axle already described, except that the axle is divided into two parts of different thickness, and the rope is fastened by both ends to the axle. The weight to be lifted is attached to a movable pulley suspended in the bight of the rope. When the wheel turns one way, the rope winds on the larger axle and unwinds from the smaller one; when it turns the other way, the reverse happens. Represent the radius of the larger axle by r, that of the smaller axle by s, and the distance of the handle from the axis by R as before. Then when the axle turns once, a length of rope equal to $2\pi r$ is wound up on the larger axle and a length equal to $2\pi s$ is unwound from the smaller one. The weight W is therefore lifted through the distance $\pi (r - s)$. Setting the quantities of work done equal, we have $2\pi RP = \pi (r - s) W$, so that the mechanical advantage of the system is $\frac{W}{P} = \frac{2R}{r - s}$.

The Parallelogram of Forces

38. Nonparallel Forces. — In our examination of the fundamental case of equilibrium about an axis (§ 19) we found that two forces will make equilibrium when their moments about the axis are equal and opposite in sense, whether the forces are parallel or not. Evidently in such cases the axis is acted on by a force which is the resultant of the applied forces, and the reaction of the axis is their equilibrant. We have found the equilibrant when the applied forces are parallel, and the question now arises of finding the equilibrant in the more general case, when the applied forces are not parallel.

The same question is also suggested by the solution of the problem of the inclined plane, for obviously the two forces which have been specified as acting on the body on the plane would not of themselves be sufficient to cause equilibrium. There must be a third force, arising from the reaction of the plane, which is the equilibrant of the other two.

The answer to this question may be obtained by the use of the principle of moments or of the principle of work; but it is most satisfactory to obtain it directly by experiment, reserving for further study the relation of the new principle which we shall establish to those already considered.

39. Equilibrant of Two Nonparallel Forces. — When the lines of two nonparallel forces lie in one plane, the line of their equilibrant lies in the same plane, and the lines of all three forces meet at a common point.

This proposition may be demonstrated by the following argument:

If an axis is passed through the point at which the lines of the two given forces intersect (Fig. 30), it is plain that neither of the forces will produce rotation about the axis; and therefore their equilibrant will not produce rotation about the axis, and so its line will pass through the point of intersection of the given forces. The lines of the three forces therefore meet at a common point.

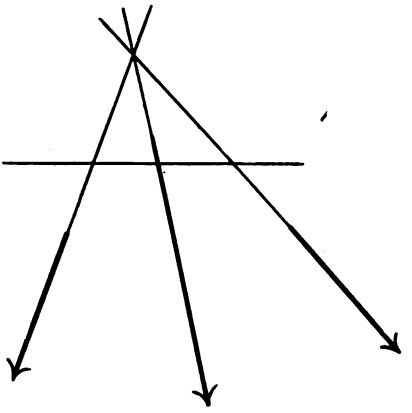


Fig. 30.

Similarly, if any axis is drawn in the plane of the given forces and cutting their lines, it is plain that neither of the forces will produce rotation about the axis; and therefore their equilibrant will not produce rotation about the axis, and so its line will pass through the axis. The lines of the three forces therefore lie in a common plane.

40. The Parallelogram of Forces. — By the help of the proposition which has just been given we can simplify considerably the fundamental experiment by which we determine the law connecting the equilibrant with the two given forces. Instead of setting up a body similar to that used in our experiments on moments, and applying an equilibrating force to its axis, we may simply apply two forces to a common point and determine the single force which will make equilibrium with them when applied at the same point.

A weight hung by a cord may be suspended from a point on another cord whose ends are attached to two spring balances. Each of the balances will indicate a force; and, in successive experiments with different weights and with different points of attachment of the vertical cord, the magnitudes of the three forces and their directions may be noted. We may then deduce the desired relation among the forces by a study of these observations.

In giving the results of observation in the form of a general rule it is not only customary but more generally useful to state the relation obtained as the relation of the two given forces to their resultant rather than to their equilibrant. The rule therefore becomes a rule for finding the resultant of two nonparallel forces. The examination of observations shows that they may always be represented by the following construction (Fig. 31):

From any point as origin draw two lines in the directions of the given forces; on them measure off lengths numerically equal, on some convenient scale, to the magnitudes of the forces; complete a parallelogram on these two lengths as sides, and draw its diagonal from the origin. The direction of this diagonal is the direction of the resultant, and its length is numerically equal on the same scale to the magnitude of the resultant.

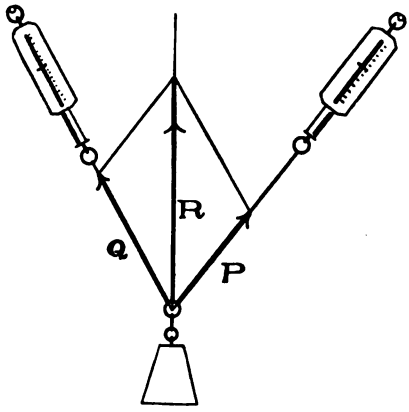


Fig. 31.

This rule or law is known as the law of the *parallelogram of forces.*

Special cases of it were cited by Aristotle (384–322 B.C.) and by Stevinus (1605). It was first clearly stated by Newton (1687), and it was independently discovered, and its application to statics developed, by Varignon (1725). The way in which Newton derived it will be given later.

If we make the convention that a line drawn in the proper direction and of the proper length *represents* a force, we may state the parallelogram law more concisely by saying:

If two forces are represented by the sides of a parallelogram drawn from one of its angular points as origin, the diagonal drawn from the same point represents their resultant.

The process of finding the resultant of forces is often called *compounding* the forces.

41. Composition of Forces. — If several forces are given, their resultant may be found by compounding two of the forces to find their resultant, using this resultant with another of the forces to find a new resultant, and proceeding in this way until all the forces have been used and a final resultant is reached

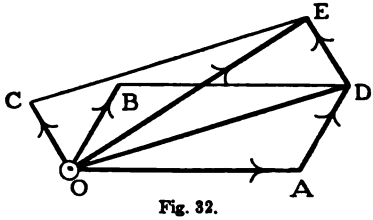


Fig. 32.

A simpler way to accomplish the same result may be illustrated by considering the case of three given forces. In Fig. 32 let the lines $\overline{OA}$, $\overline{OB}$, $\overline{OC}$ represent three forces, of which the resultant is desired. We compound $\overline{OA}$ with $\overline{OB}$,

obtaining the resultant $\overline{OD}$, and compounding $\overline{OC}$ with $\overline{OD}$ we obtain the final resultant $\overline{OE}$. But inspection of the construction shows us at once that the line $\overline{AD}$ is equal and parallel to $\overline{OB}$, and the line $\overline{DE}$ equal and parallel to $\overline{OC}$, so that the final resultant may be obtained by drawing from O the line $\overline{OA}$ representing the first force, from A the line $\overline{AD}$ representing the second force, and from D the line $\overline{DE}$ representing the third force. The line $\overline{OE}$ which is drawn from the origin to the end of the last representative line, represents the resultant.

This illustration may obviously be expanded by introducing other forces. We arrive at the following rule for the composition of forces:

From any point as origin draw a line to represent one of the forces, from the end point of that line draw a line to represent another of the forces, and proceed in this way until all the forces have been introduced. The line which may then be drawn from the origin to the end point of the last line drawn represents the resultant of the forces. This construction applies equally well to the case in which the forces are not all in one plane.

The lines representing the forces form the sides of a polygon (Fig. 33), and the construction is called the polygon of forces. If

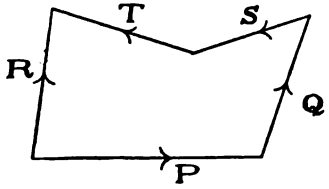


Fig. 33.

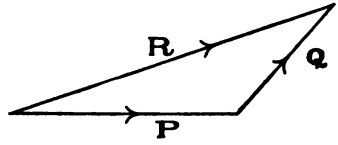


Fig. 34.

the end point of the line representing the last force taken meets the origin, so that the polygon is closed with the forces of the system, there is no resultant. The system of forces is in equilibrium.

When two forces only are considered, the lines representing the forces and their resultant form a triangle; and the construction is called the triangle of forces (Fig. 34). The lines representing three forces in equilibrium form a closed triangle.

42. Resolution of Forces. — It is often important to find two or more forces which have a resultant equal to a given force. This cannot be done unless a sufficient number of conditions are given to which the forces to be found shall conform. In the simplest case, in which two forces are to be found whose resultant is equal to a given force, two conditions determining either the direction or the magnitude of one or both forces must be given.

The process of finding these forces is called the *resolution* of a force. The forces which are found are called the *components* of the given force, and the force whose components are found is said to be resolved into them.

As an illustration of the process of resolving a force into two components we may take the case of a given force whose components in two given directions are desired (Fig. 35). From any

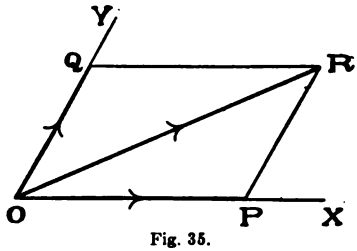


Fig. 35.

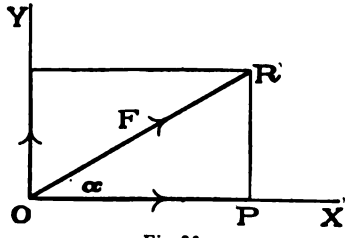


Fig. 36.

point O draw a line $\overline{OR}$ representing the given force; from the same point draw two lines OX, OY of indefinite length in the given directions, which must be on either side of the direction of the given force; with the line representing the given force as diagonal, construct the parallelogram whose sides are in the given directions; the components desired are then given by the sides $\overline{OP}$, $\overline{OQ}$ of this parallelogram.

We are often asked to find the component of a force in a given direction or making a given angle with the given force. The problem as thus stated lacks one of the two conditions necessary to enable us to reach a definite result; but when it is stated in this way it is always assumed that the other component, about which nothing is said, is to be taken in a direction at right angles to the one which is desired. With this understanding, the required component can be found by drawing the line $\overline{OR}$ (Fig. 36) representing the given force, and then drawing two lines from the same origin, one of which, OX, is in the given direction and the other at right angles to it. When the rectangular parallelogram is constructed, on these two directions, about the line representing the given force as diagonal, the side $\overline{OP}$ of the parallelogram, which is in the given direction, is the component desired.

The line $\overline{OP}$, which is the intercept between the two perpendiculars on the given direction OX from the ends of the line F, is often called the projection of F on the direction OX. The component of a force in a given direction when no further specifications

are given may therefore, by analogy, be defined as the projection of the force in that direction.

If the angle between the direction of the given force F and the direction of the desired component is α, the magnitude of the component is $F \cos \alpha$.

43. Analytical Method of Compounding Forces. — The most satisfactory way to obtain the resultant of a number of forces which are specified by reference to some system of coördinate axes is to resolve each of the forces into its components along the axes, add the components of all the forces along each axis, and compound the mutually perpendicular forces thus obtained to obtain the resultant. That this method will yield the resultant desired is evident on the inspection of the simple case of the composition of two forces. For in Fig. 37, representing a triangle of forces, the algebraic sums of the projections of the two component forces upon the axes are evidently equal to the projections of the resultant upon the same axes.

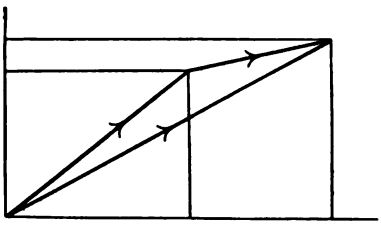


Fig. 37.

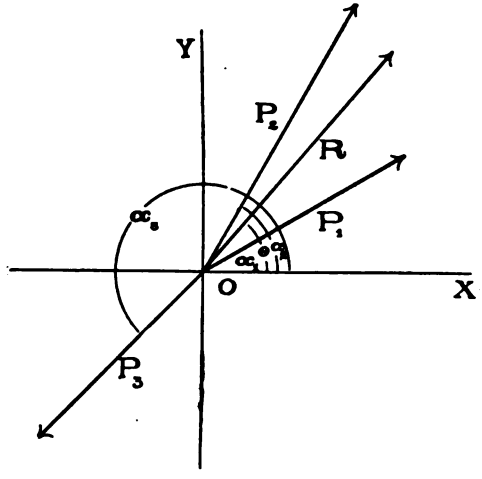


Fig. 38.

If the forces are all in one plane, say the xy-plane (Fig. 38), and are specified by giving for each its magnitude P, and the angle α which its positive direction makes with the positive direction of

the x-axis, the sums of the components of each force along the x- and y-axes are

$$\left.\begin{aligned} X &= \Sigma P\cos\alpha, \\ Y &= \Sigma P\sin\alpha. \end{aligned}\right\} \tag{13}$$

The resultant is

$$R = \sqrt{X^2 + Y^2}, \tag{13a}$$

and the angle θ which its direction makes with the x-axis is given by

$$\frac{Y}{X} = \tan\theta. \tag{13b}$$

If the forces are not in one plane, and are specified by giving for each its magnitude P and the angles α, β, γ which its positive direction makes with the three rectangular axes of x, y, z, the sums of the components of each force along the three axes are

$$\left.\begin{aligned} X &= \Sigma P\cos\alpha, \\ Y &= \Sigma P\cos\beta, \\ Z &= \Sigma P\cos\gamma. \end{aligned}\right\} \tag{14}$$

The resultant is

$$R = \sqrt{X^2 + Y^2 + Z^2} \tag{14a}$$

and the angles θ, ϕ, ψ made by its direction with the three axes are given by

$$\cos\theta = \frac{X}{R}, \quad \cos\phi = \frac{Y}{R}, \quad \cos\psi = \frac{Z}{R}. \tag{14b}$$

44. Resultant Moment of a System of Nonparallel Forces. — To show that the parallelogram of forces is in harmony with the principle of moments, we need only to show that the sum of the moments about an axis of two or more nonparallel forces is equal to the moment of their resultant about the same axis. We shall consider first the case in which the axis is perpendicular to the plane of the forces.

We begin by interpreting the formula for the moment of a force about an axis perpendicular to the plane in which the force lies in a way that will be useful in proving the desired relation. By the definition of a moment of force the moment of the force P (Fig. 39), applied at the point A, about an axis at O, perpendicular to the plane of the paper, in which the force lies, is equal to Pp. Now draw the line l connecting the points O and A. Represent by α the angle between the lines l and p. Then $p = l\cos\alpha$ and the moment of the force may be represented by $Pl\cos\alpha$. But $P\cos\alpha$ is the component of the force P perpendicular to the line l; for the triangle of forces ABC is similar to the triangle OAD, and the angle CAB is

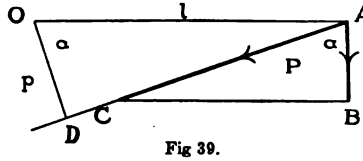


Fig 39.

equal to the angle α. Therefore the moment of the force P about the axis is the product of l and of $P\cos\alpha$, and in general the moment of a force about an axis may be defined as the product of a line drawn from the axis to any point in the line of the force and the component of the force perpendicular to that line.

Now let us consider two forces and any axis through O perpendicular to their plane (Fig. 40). The sum of their moments about the axis is the product of the line l drawn from O to A, the common point of the lines of the two forces, and the sum of their components perpendicular to that line. But the sum of these components is equal to the component of the resultant of the two forces perpendicular to the same line, as appears on inspection of the parallelogram by which the resultant is obtained. Therefore the resultant of any two forces, as given by the parallelogram law, has a moment about any axis perpendicular to the plane of the forces which is equal to the sum of the moments of the forces about the same axis.

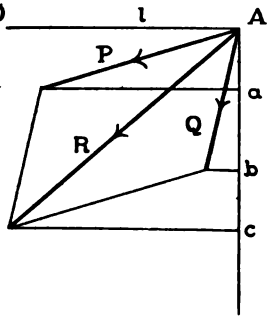


Fig 40.

Since the equilibrant has a moment about any point equal and opposite in sense to that of the resultant, the sum of the moments of three nonparallel forces in equilibrium about any axis perpendicular to the plane of the forces is equal to zero. Manifestly a similar demonstration will apply if more than three forces are taken into account.

In the more general case in which the axis about which moments are taken is oblique to the plane of the forces, it is necessary to give a definition of the moment of a force about an oblique axis. If such a force is resolved into its components parallel with the axis and in a plane perpendicular to the axis, it is clear that the component parallel with the axis can have no effect with respect to rotation about the axis, and in considering the effective moment of the force about the axis this component may be disregarded. We therefore define the moment of a force about an oblique axis as equal to the moment about that axis of its component in a plane perpendicular to the axis. With this definition the principle of moments may be stated generally and proved without much difficulty to be still consistent with the parallelogram of forces. We shall not, however, enter upon this question any further.

45. Work Done by a System of Nonparallel Forces. — To show that the parallelogram law is consistent with the principle of work it is sufficient to show that if two nonparallel forces act on a body which is slightly displaced, the work done by the two forces is equal to the work that would be done, for the same displacement, by their resultant. To show this we shall consider first the case of two forces so acting that the displacement of the body lies in the plane of the forces. In this case it is easy to see, by inspection of a diagram (Fig. 41) in which the forces P and Q and their resultant R are represented by the parallelogram construction, and the displacement is drawn from the origin in its proper direction, say ON, that the sum of the components $\overline{OC}$ and $\overline{OE} = \overline{CD}$ of the two forces in the direction of the displacement is equal to the component $\overline{OD}$ of their resultant in the same direction; and hence that the work done by the resultant is equal to the work done by the given forces.

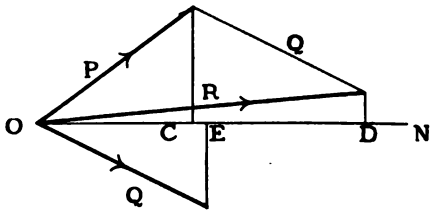


Fig. 41.

If the displacement of the body is not in the same plane as the forces, we may resolve it into two components, one of which is in the plane of the forces and the other perpendicular to that plane. This perpendicular component will not be involved in the work done by the forces, and the work done by them will depend only upon the component which lies in their plane. To this component the demonstration already given applies, and the proposition already stated holds for a displacement in any direction.

It may be that one of the two forces which are doing work on the body is so directed that the work which it does is negative. In that case, on carrying out the construction in the same way as has already been described, it will be seen (Fig. 42) that the algebraic sum of the quantities of work done by the two forces is equal to the work done by the resultant of the two forces.

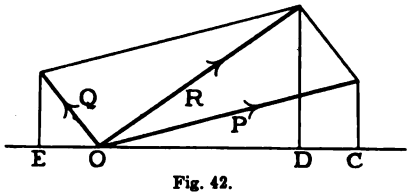


Fig. 42.

If more than two forces act, we may proceed as in the above demonstration with two of the forces, and repeat the demonstration for the resultant of these forces and a third force, and so

on until all the forces have been considered. In this way we may demonstrate the general theorem that the work done by any number of forces is equal to the work done by their resultant.

EXAMPLES, III

1. *To deduce the law of equilibrium on an inclined plane by the use of the parallelogram of forces.*

Resolve the weight Q, which rests on the plane (Fig. 43), into its components parallel and normal to the plane. The angle α made by the normal component with the line of the weight is equal to the angle of inclination of the plane. The normal component, $Q\cos\alpha$, is equilibrated by the reaction of the plane. The parallel component, $Q\sin\alpha$, is directed down the plane, and must be equilibrated by a force $P = Q\sin\alpha$, if the body is to be kept in equilibrium.

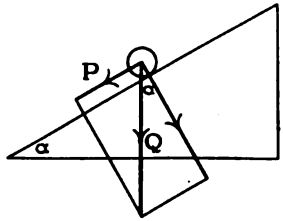


Fig. 43.

In a similar manner we may deduce the formula for the case in which the applied force is parallel to the base of the plane.

2. *Three forces, A, B, C, are in equilibrium; the angle between the lines of B and C is α; of A and C is β; of A and B is γ. Show that*

$$\frac{A}{\sin\alpha} = \frac{B}{\sin\beta} = \frac{C}{\sin\gamma}.$$ (Lami's theorem.)

This follows from the relation between the sides of triangles and the sines of the angles opposite to them.

3. *Three forces being given which are in equilibrium, to find the angles between their lines.*

Any of the formulas by which the angles of a triangle are expressed in terms of the sides may be used. For example, if the forces A, B, C are drawn as the sides of a triangle, and the angle between the lines of B and C is α, we have

$$A^2 = B^2 + C^2 + 2BC\cos\alpha;$$

and so for the other angles.

We may set $S = \dfrac{A+B+C}{2}$, form the expression

$$Q = 2\sqrt{S(S-A)(S-B)(S-C)},$$

and use the formulas $\sin\alpha = \dfrac{Q}{BC}$, $\sin\beta = \dfrac{Q}{AC}$, $\sin\gamma = \dfrac{Q}{AB}$.

The angles found in the tables from these sines will be the angles of the triangle of forces. The true angles between the lines of the forces are the complements of these, which have the same sines.

4. *A ladder of weight W and length l leans against a smooth wall at an angle α from the vertical and is prevented from slipping on an otherwise smooth floor by a cleat* (Fig. 44). *Find the reactions.*

The reactions are perpendicular to the smooth surfaces. We have $R_1 = W$, $R_2 = R_3$, and $\dfrac{Wl}{2}\sin\alpha = R_3 l\cos\alpha$, from the principle of moments, whence $R_2 = R_3 = \dfrac{W}{2}\tan\alpha$.

The resultant of R_1 and R_2 makes an angle β with the vertical given by $R_2 = R_1 \tan\beta$, $\tan\beta = \dfrac{\tan\alpha}{2}$.

These results also follow from the condition that the lines of three forces in equilibrium meet at a common point.

5. *A heavy rod is supended by two cords attached to its ends. We have given the angle α made by the rod with the vertical and the angle β made by the cord attached to the lower end of the rod with the vertical. Find the tensions in the cords*

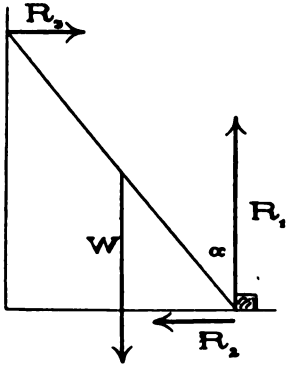


Fig. 44.

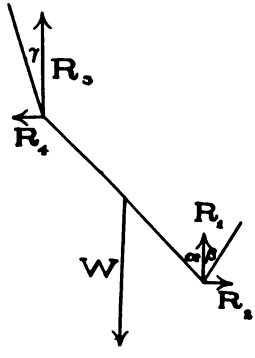


Fig. 45.

The weight of the rod is W, its length l (Fig. 45). The tensions in the cords are resolved vertically and horizontally. We have

$$W = R_1 + R_3,\ R_2 = R_4,\ R_2 = R_1 \tan\beta,\ \frac{Wl}{2}\sin\alpha + R_4 l\cos\alpha = R_4 l\sin\alpha;$$

and obtain

$$R_1 = \frac{W}{2}\,\frac{\tan\alpha}{\tan\alpha + \tan\beta},$$

$$R_2 = R_4 = \frac{W}{2}\cdot\frac{\tan\alpha\tan\beta}{\tan\alpha + \tan\beta},$$

$$R_3 = \frac{W}{2}\cdot\frac{\tan\alpha + 2\tan\beta}{\tan\alpha + \tan\beta}.$$

The angle γ made by the cord attached to the upper end of the rod with the vertical is found from $\tan\gamma = \dfrac{R_4}{R_3} = \dfrac{\tan\alpha\tan\beta}{\tan\alpha + 2\tan\beta}$.

It is of interest to discuss the changes in the component tensions as the angle β approaches 90°; also as the angle α approaches 90°.

6. *A heavy rod is suspended by a cord from its upper end, and rests with its lower end on a smooth floor (or on a board floating in water), making an angle α with the vertical. Find the tension in the cord and the reaction at the floor.*

In this case, from the given conditions, $R_2 = R_4 = 0$ (Example 5). From the results of that example, we then get $\tan\beta = 0$, $\tan\gamma = 0$, so that the cord is vertical, as well as the reaction at the floor. We have further $R_1 = R_3 = \dfrac{W}{2}$.

7. *A door swings on two hinges, supported by the lower one; find the reactions at the hinges.*

There is no vertical reaction at the upper hinge; call the horizontal reaction there R_3, the vertical and horizontal reactions at the lower hinge R_1 and R_2 respectively. Call the distance between the hinges l, the distance of the center of gravity of the door from the line of the hinges d, and the weight of the door W. The use of the principle of moments gives for equilibrium about the lower hinge, $R_3 l = Wd$; also $R_1 = W$, $R_2 = R_3$.

Static Friction

46. Friction. — In the cases of equilibrium which we have used to illustrate the application of the principles of statics the bodies were often considered as fixed at some point or at different points, and as acted on at those points by certain forces called reactions. The way in which the body was supposed to be fixed was such that no matter what the magnitude of the force applied at the fixed point might be, the reaction at the point would remain equal to it, so that equilibrium would be maintained. In our experience, however, there occur numerous examples of another type of reaction in which equilibrium fails if the applied force passes a certain limit. These cases are those in which the reaction is due to friction.

Without treating the subject of friction generally at this time, it is convenient to consider it sufficiently to enable us to deal with problems in statics in which frictional forces occur. Such problems are often of great practical importance, and though the laws of friction do not represent the behavior of rough bodies closer than to a first approximation, yet even thus the solutions obtained from them represent a much closer approximation to real conditions than the solutions of abstract mechanics.

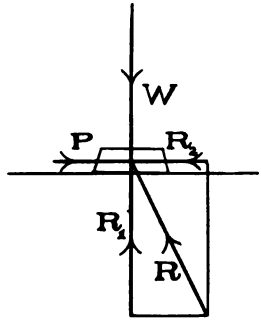


Fig. 46.

When a body is pressed against the surface of another body by any force, experience proves that a force applied to it parallel to the surface against which it is pressed encounters a reaction. This reaction is to be conceived of as a force applied to the body by the surface against which it is pressed. The case considered is illustrated in Fig. 46. The body is pressed against the plane surface by the normal force W and is pushed parallel to the plane by the force P. The total reaction acting on the body is a force R, which is the equilibrant of the forces W and P; this can be resolved into components R_1 and R_2, which equilibrate W and P respectively.

47. Coefficient of Friction. — Experience further shows that when the force P exceeds a certain magnitude the reaction R_2 fails to equilibrate it, and the body starts to move. The magnitude of the tangential force which starts the body in motion over the surface of another one is found by experiment to depend on the nature of the materials constituting the bodies, on the irregularities of their surfaces, and on the force with which the surfaces are pressed together.

In any given case in which experiments are tried with a particular body pressed against a particular surface it is found that to a first approximation the force P which will just break equilibrium and start the body in motion is proportional to the force W with which the body is pressed against the surface. This proportionality is generally expressed by the formula

$$P = \mu W. \tag{15}$$

The constant μ is the factor of proportion; it is called the *coefficient of static friction*. Its value is independent of the extent of surface over which the bodies meet, but it has different values for different bodies, depending upon the character of the surfaces in contact.

48. Angle of Friction. — It is convenient to represent the coefficient of friction as the tangent of a certain angle. To do so, let us consider the diagram of the reactions called forth by the forces acting on the body (Fig. 47). At the instant of motion the force P will have attained a certain value, given by $P = \mu W$; the resultant of P and W will make a certain angle ϵ with the normal to the surface; and from the parallelogram we shall have $P = W \tan\epsilon$. Comparing the two expressions for P, we have $\mu = \tan\epsilon$, and see that we may express the coefficient of static friction by the tangent of the angle between the resultant of the forces applied to the body at the instant of motion and the normal to the surface over which it moves. This angle is called the *limiting angle* or the *angle of friction*.

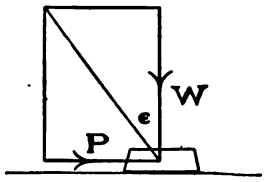


Fig. 47.

In ideal problems bodies in contact are often supposed to be perfectly smooth or frictionless. In such a case the coefficient of friction vanishes, and the angle of friction becomes zero, so that the reaction of a smooth surface on a body is always normal to the surface. In actual arrangements this condition is never attained, but highly polished bodies will sometimes show a very small coefficient of friction.

EXAMPLES, IV

1. *The coefficient of friction of a body placed on an inclined plane may be found by determining the angle of inclination of the plane at which the body begins to slip.*

Resolve the weight W of the body, which is a force normal to the base of the plane, into its components normal and parallel to the face of the plane. The normal component $W \cos \alpha$ presses the body against the plane; the parallel component $W \sin \alpha$ pushes it down the plane. For the value of α at which the body slips we have $W \sin \alpha = \mu W \cos \alpha$, or $\mu = \tan \epsilon = \tan \alpha$.

The angle α thus found is therefore the angle of friction for this case.

2. *A weight W on an inclined plane is attached by a cord going over a pulley at the top of the plane to another weight P which hangs vertically. For a certain angle of inclination α_1 of the plane the weight on it will slip down the plane; for a certain other angle α_2 it will slip up the plane; find the angles and show that $\alpha_1 - \alpha_2 = 2\epsilon$.*

The frictional reaction always acts against the motion, so that when the weight slips downwards the reaction is in the same direction as P, when it slips upwards the direction of the reaction is reversed. The equations of equilibrium at the instant of slipping therefore are

$$W \sin \alpha_1 = P + \mu W \cos \alpha_1,$$
$$W \sin \alpha_2 = P - \mu W \cos \alpha_2.$$

These equations determine the angles.

Eliminating P, we have

$$\sin \alpha_1 - \sin \alpha_2 = \mu (\cos \alpha_1 + \cos \alpha_2).$$

Replacing μ by $\tan \epsilon = \dfrac{\sin \epsilon}{\cos \epsilon}$ and transposing terms we obtain

$$\sin \alpha_1 \cos \epsilon - \cos \alpha_1 \sin \epsilon = \sin \alpha_2 \cos \epsilon + \cos \alpha_2 \sin \epsilon,$$
$$\sin (\alpha_1 - \epsilon) = \sin (\alpha_2 + \epsilon),$$
$$\alpha_1 - \alpha_2 = 2\epsilon.$$

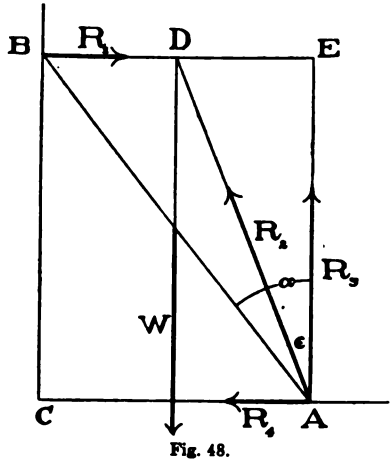


Fig. 48.

3. *A ladder stands on a floor and leans against a wall. Assuming that the floor is rough and the wall smooth, find the greatest inclination which the ladder can have without slipping.*

The weight of the ladder (Fig. 48) is a force W applied at its center. The other forces which act on the ladder are a normal reaction R_1 at B from the smooth wall and an inclined reaction R_2 at A from the rough floor; this may be resolved into a normal component R_3 and a horizontal component R_4 due to the friction.

Using the principle of moments we have $R_1 = R_4$, $W = R_3$, and, taking moments about the point A, representing the length of the ladder by l and the angle of inclination from the normal to the floor by α, $R_1 l \cos \alpha = W \dfrac{l}{2} \sin \alpha$. Also just at the point of slipping $R_4 = \mu R_3$.

From these four equations the three unknown reactions and the angle α may be obtained as follows:

$$R_1 = R_4 = \mu W, \quad R_3 = W, \quad \tan \alpha = 2\mu.$$

We may get the value of α very simply by using the theorem that three forces in equilibrium meet at a common point. The line of the normal reaction at B and the line of the weight meet at D, and therefore the line of the inclined reaction at A must also pass through D. At the instant of slipping the angle between this inclined reaction and the normal is the angle of friction ϵ. From the figure we have plainly $BE = 2DE$; and $BE = AE \tan \alpha$, $DE = AE \tan \epsilon$. $\therefore \tan \alpha = 2 \tan \epsilon$.

4. *A man ascends a ladder standing as in Example* 3, *at a given angle of inclination* α; *how high can he go before the ladder slips, assuming that the weight of the ladder may be disregarded?*

Let d represent the distance of the man from the bottom of the ladder and P his weight, the other magnitudes being designated as in Example 3. The equations then are

$R_1 = R_4$, $R_3 = P$, and, taking moments around A, $R_1 l \cos \alpha = Pd \sin \alpha$. The condition of slipping gives $R_4 = \mu R_3$.

In this example α is given, and d is the quantity to be determined. We obtain

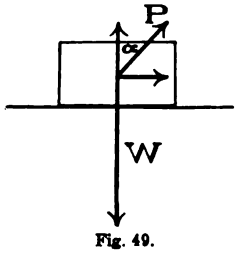


Fig. 49.

$$R_1 = R_4 = \mu P;\ R_3 = P;\ d \tan \alpha = \mu l = l \tan \epsilon.$$

The height to which the man can ascend before the ladder slips is less than the length l if $\alpha > \epsilon$. If $\alpha < \epsilon$ he can ascend to the top without the ladder's slipping.

5. *A heavy weight* W *rests on a rough floor, and a force* P *is applied to it by a cord; if the force cannot start the weight moving when directed along the floor, find whether it can do so when directed at any other angle, and if so, find the value of the angle.*

When the cord is inclined, a component of P acts upwards (Fig. 49) and diminishes the force with which the weight is pressed against the floor. We have

$$\mu (W - P \cos \alpha) = P \sin \alpha,$$

$$W - P \cos \alpha = P \sin \alpha \frac{\cos \epsilon}{\sin \epsilon},$$

$$W \sin \epsilon = P (\sin \alpha \cos \epsilon + \sin \epsilon \cos \alpha),$$

$$W \sin \epsilon = P \sin (\alpha + \epsilon),$$

from which the value of α can be obtained in any specific case. If the values of P, W, and ϵ are such that the equation cannot be satisfied by any value of α, i.e. if $P < W \sin \epsilon$, the force cannot move the weight. The least force that can move the weight is that given by the value of P when $\sin (\alpha + \epsilon) = 1$, or when $\alpha = \frac{\pi}{2} - \epsilon$.

6. *Explain in general terms why a screw can turn and move forward in a nut but cannot be drawn out of it by a direct pull.*

In the case of the direct pull the component of the pull parallel to the thread of the screw is so small that it does not equal the product of the coefficient of friction and the component perpendicular to the thread, so that the screw cannot turn. In the other case the component force applied to the thread of the screw is greater than the same product, and the screw can turn.

CHAPTER III

DYNAMICS

49. The Problem of Dynamics. — In very many cases the result of applying a force to a body is that the body is either set in motion, or has its motion changed in some way. The science of *dynamics* deals with the relations between these motions and the forces which occasion them. The general problem of dynamics is to find the forces when the motions of the bodies are given, or to find the motions when the forces are given.

DYNAMICS OF A PARTICLE

Force and Motion

50. Motion Due to Constant Force. — When a body is not sustained by a force acting upwards it always falls toward the earth at a rate which evidently increases as it nears the earth. The motion of such a body is one of the simplest exhibitions of motion caused by a force. It attracted the attention of Galileo, and the study of it by him was the first step taken in the direction of enlarging the subject of mechanics to include the study of motions of bodies as well as the study of cases of equilibrium. Galileo succeeded in establishing the law connecting the motion of a body with the force acting on it, when the force is the weight of the body. Following his course of thought we begin with certain preliminary notions connected with motion in general

51. Velocity. — When a point moves in a straight line in such a way that it passes over equal distances in any arbitrarily chosen equal times, or when a point moves in a straight line in such a way that the ratio of the distance passed over to the time occupied by the point in passing over that distance is a constant, the point is said to have a *constant velocity*. When these two conditions, of motion in a straight line and of a constant ratio between the space and the time, are not fulfilled, or when either of them is not fulfilled, the point is said to have a *variable velocity*.

When the motion is considered without reference to the shape of the path, the ratio between the space and the time is often called

the *speed* of the moving point. The speed may also be constant or variable.

The value of the variable velocity, at any point in the line or path over which the point moves, may be defined by supposing that the moving point, after it passes through the point on the line at which its velocity is desired, moves on with a constant velocity, equal to that which it has as it passes through the point on the line. This constant velocity will be the velocity at the point on the line.

The velocity at a point may be defined generally as the limit of the ratio of the distance traversed from that point to the time taken to traverse it, as the time, and so also the distance, approach zero as a limit. To appreciate this definition, we consider a point on the line of motion at which we wish to determine the velocity. If we measure off from that point any distance along the line and divide it by the time taken by the moving point to traverse that distance, the ratio thus obtained may be called the *average velocity* of the moving point during that time. By taking shorter and shorter distances from the point, the ratios obtained may be made to approach as nearly as we please to the velocity at the given point, and in the limit the ratio of the infinitesimal distance traversed to the infinitesimal time taken to traverse it will be the velocity at the given point. If the ratios thus obtained are all equal and the path is straight, the velocity is constant; if they differ or if the path is curved, the velocity is variable. The direction of the tangent to the path at the given point is the direction of the desired velocity.

In the case of constant velocity, if we represent by s the distance which the moving point traverses, and by t the elapsed time, the velocity v is given by the formula

$$v = \frac{s}{t}. \qquad (16)$$

This formula also represents the average velocity of a point moving with a variable velocity over the distance s in the time t.

The unit of velocity is the velocity of a point moving with constant velocity which traverses a unit distance in unit time. In the c.g.s. system it is the velocity of a point moving with constant velocity which traverses one centimeter in one second.

52. Acceleration. — When a point moves with a variable velocity it is said to have *acceleration*. If the point moves in a straight

line and in such a way that its velocity changes by equal amounts in any arbitrarily chosen equal times, its acceleration is said to be constant. When these two conditions are not fulfilled, or when either of them is not fulfilled, the acceleration of the point is variable.

In the case of constant acceleration, its numerical value is found by dividing the change in velocity which occurs in any arbitrarily chosen time by that time, the quotient obtained being the desired acceleration. If we represent acceleration by a, the velocity of the point at the beginning of the time t by v_0, and that at the end of the time t by v, the formula defining and measuring a constant acceleration is

$$a = \frac{v - v_0}{t}. \tag{17}$$

According as the change in velocity is an increase or a decrease, the acceleration is positive or negative.

The unit of acceleration is the acceleration of a point whose velocity increases uniformly by one unit in unit time. In the c.g.s. system it is the acceleration of a point whose velocity increases in one second by the velocity of one centimeter per second. It is often called an acceleration of one centimeter per second per second.

53. **Effect of a Constant Force.** — Heavy bodies evidently fall in straight lines and with increasing velocity. Their motion is therefore accelerated. The problem taken up by Galileo was the determination of the law of their acceleration. In attacking it he used the method of hypothesis and deduction described in § 2.

Following his line of argument, we assume, as a hypothesis, that the velocities of a falling body at different instants are proportional to the times during which it has been falling, or that a falling body has a constant acceleration. We then proceed as follows to establish a consequence of this hypothesis which can be tested by experiment.

When a body that has been held above the earth is first released, it starts from rest and its initial velocity is zero. After the lapse of t seconds its velocity is v. In the t seconds it will traverse a distance s which is equal to the distance which it would traverse in t seconds if it were moving with a constant velocity equal to the average of the velocities with which it actually moves. On the hypothesis that its acceleration is constant and therefore that its velocity has been increasing at a uniform rate, this average velocity

is $\frac{v}{2}$, and the distance which it will traverse in the t seconds is therefore

$$s = \frac{vt}{2} = \frac{1}{2}at^2, \tag{18}$$

since $v = at$.

Galileo illustrated this way of averaging the uniformly increasing velocity by the accompanying diagram (Fig. 50), which needs no explanation.

That this conclusion is correct may be seen from the following considerations: Let us suppose that the change in velocity takes place not uniformly but by small equal increments occurring at the ends of small equal intervals of time. Construct a diagram (Fig. 51) by marking off along a horizontal line equal distances Aa, ab, bc, . . . representing equal increments of time τ, and by drawing, from the points thus determined, vertical lines representing

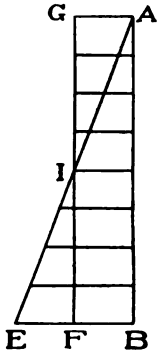


Fig. 50.

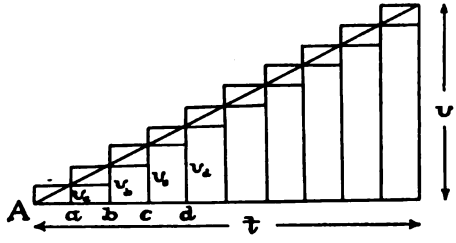


Fig. 51.

the velocities which the point will assume at the corresponding instants. On the lines thus determined construct parallelograms. The area of each of these parallelograms will be numerically equal to the product of a time interval and of the velocity v_a, v_b, v_c, . . . which the point possesses during that interval, and will therefore be the distance $s_a = v_a\tau$, $s_b = v_b\tau$, $s_c = v_c\tau$, . . . which is traversed by the moving point during that interval. The sum of all these areas is therefore equal to the entire distance $s = s_a + s_b + s_c + \ldots$ traversed by the moving point in the time t under the given conditions.

If the changes in velocity occur at the beginnings of the time intervals, this sum is the area of the outside figure; if at the ends, the area of the inside figure. As the time intervals are taken shorter and shorter, these two areas become more nearly equal. In the limit, as the time intervals and the associated changes of velocity become infinitesimal, so that the velocity changes uniformly with the time, these areas become equal and equal to the area of the triangle whose base is the line $t = \Sigma\tau$ and whose altitude is the velocity v of the point at the end of the time t. The distance traversed by the moving point is represented numerically by the area of the triangle, and since this area is equal to half the base t multiplied by the altitude v of the triangle, we reach the formula for the distance traversed which has already been given.

This relation between the distance traversed by the falling body and the elapsed time may be tested by experiment, and according as the results of experiment exhibit this relation or not, the hypothesis from which it was deduced is confirmed or disproved. This test is now often made directly, and the proportionality between the distance traversed and the square of the time has been abundantly demonstrated by observations on freely falling bodies.

It was not possible for Galileo to do this, chiefly because he possessed no proper means for the accurate measurement of short intervals of time. He was compelled to adopt a method in which the essentials of the experiment were preserved, while the time elapsing during the motion was sufficiently increased to make its measurement possible without excessive error. When a body is on an inclined plane, the force which draws it downward on the plane is a component of its weight, which is constant so long as the inclination of the plane is not changed; and it is reasonable to suppose that its effect on the body will follow the same law as that produced by the whole weight when the body falls freely. Galileo therefore adopted the plan of allowing a smooth brass ball to roll down an inclined plane and determining the times taken by the ball to traverse different distances. He measured the time by means of a large vessel, filled with water, in the bottom of which was a small opening. This opening was stopped with the finger until the ball was released. At the moment the ball was released, the finger was removed, and the water allowed to flow out into a small cup until the ball passed a marked point on the plane, when the opening was closed again. The water that flowed into the cup was weighed, and the time which had elapsed was taken proportional to its weight. By making a large number of experiments of this kind, to eliminate experimental errors in the final averages taken, Galileo was able to show that the distances passed over by the ball after starting from rest were proportional to the squares of the times during which it was moving.

The results of these experiments are therefore consistent with the hypothesis that a constant force proportional to the weight of a body, and acting on it, gives to it a constant acceleration; and lead by an easy inference to the special conclusion that a freely falling body moves with a constant acceleration.

The production of a constant acceleration by a constant force is illustrated by the use of an apparatus called the Atwood's machine. This consists essentially of a light and easily running

pulley, over which is passed a cord sustaining two equal weights (Fig. 52). This system is in equilibrium, and if the weights are set moving by a mere push or impulse so as to turn the pulley either way, the motion is appreciably uniform through the range permitted by the length of the cord. The pair of weights, therefore, may be considered as a body at rest and freed from the action of its own weight, at least in so far as its weight is a force which sets it in motion. If a small additional body, called an overweight, is placed on one of the weights, the system will start moving under its action, and the movements of the system can be examined. By determining the times the system takes to move from rest to a platform placed at various points, it can be shown that the distances traversed from rest are proportional to the squares of the times. By arranging a ring so as to pick off the overweight after the system has been moving for a known time, and then determining the times the system takes to move from the ring to the platform placed at various distances below it, it can also be shown that the velocities acquired in the times during which the overweight has acted are proportional to those times. With this machine the law that a constant force produces a constant acceleration is verified both indirectly by the illustration of the formula $s = \frac{1}{2}at^2$, and more directly by the illustration of the formula $v = at$.

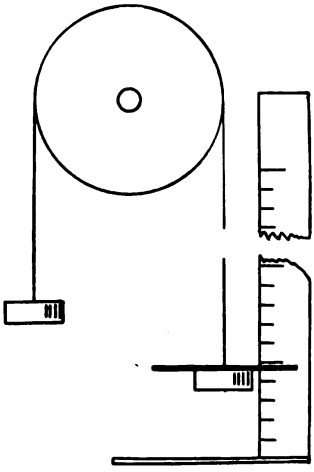


Fig. 52.

Both Galileo's experiment and the experiment with the Atwood's machine are open to criticism. In Galileo's experiment the ball rolls instead of moving without turning, and we must inquire what effect its rolling will have on its linear motion. Similarly, in the Atwood's machine, the pulley is set turning by the moving system of weights, and its influence on the motion must be studied. We cannot take up the inquiry at present; but it will be shown further on (Examples IX, 2, 3) that these experiments are correctly interpreted as illustrations of the effect of a constant force in giving a body a constant acceleration, and that the influence of the rolling in the one case, and the turning of the pulley in the other, is shown in diminishing the magnitude of the acceleration.

54. Motion with Constant Acceleration. — The laws of falling bodies, or the laws of motion of any body which starts from rest

and moves with a constant acceleration, are given by the following formulas:

$$\left.\begin{aligned} v &= at, \\ s &= \tfrac{1}{2}at^2, \\ v^2 &= 2as. \end{aligned}\right\} \tag{19}$$

In these formulas t represents the time interval which elapses from the instant at which the body starts to move to that at which the measurement of velocity or distance is made; s represents the distance traversed in the time t, and v represents the velocity possessed by the body at the end of the time t, after it has traversed the distance s. The last formula is obtained from the other two by the elimination of t.

In case the body is already moving with the initial velocity u when it enters the space s at the beginning of the time t, this initial velocity will be added to that acquired by the body by reason of its acceleration, and the final velocity v at the end of the time t is given by

$$v = u + at. \tag{20a}$$

The average velocity during the time t will be $\frac{u+v}{2}$, and the space traversed will be therefore

$$s = \frac{u+v}{2}t,$$

or

$$s = ut + \tfrac{1}{2}at^2. \tag{20b}$$

From these equations we have also

$$v^2 - u^2 = 2as. \tag{20c}$$

These equations hold whether the acceleration is positive or negative.

55. Acceleration of Falling Bodies. — Our experiments on the laws of falling bodies are much simplified because of the remarkable fact, discovered by Galileo, that neither the material constituting the body, nor its weight, affects the acceleration with which it falls. He showed that all falling bodies which are heavy enough to move through the air, without their motion being seriously affected by its resistance, have the same acceleration. He did this by allowing different bodies to fall from the top of the Leaning Tower of Pisa, releasing them at the same instant and observing that they reached the ground together. We can confirm his conclusion by similar experiments, though the final confirmation of it requires more elaborate methods; and by operating in a vacuum we can show that the

law which holds for heavy bodies holds also for light bodies when the resistance of the air does not affect their motion.

We give to this acceleration, which is common to all falling bodies, the symbol g. It is called the *acceleration of gravity*, or often simply *gravity*. Its value is slightly different at different places, but for all ordinary purposes it may be taken as equal to 980 units of acceleration, in the c.g.s. system, or to 32 in feet and seconds. The method by which this value is determined will be given later.

56. Motion on an Inclined Plane. — Galileo deduced some propositions about motion down inclined planes, which can be tested by experiment. In every case the results of experiment confirm the initial hypothesis that a constant force produces a constant acceleration.

When a body moves down an inclined plane, the force which moves it is the component of its weight which is parallel to the plane. We assume, as Galileo did, that its acceleration on the plane will be to its acceleration if it is falling freely as this component is to its whole weight. If we represent by p the height of the plane, by s the length of the plane, by a its acceleration on the plane and by g its acceleration when falling freely, this proportion may be expressed by the equation $as = gp$.

Starting with this relation we may prove the following theorem: The times of descent down inclined planes of the same height are proportional to their lengths. For if the acceleration on the plane is constant we have $s = \frac{1}{2}at^2$ and hence $s^2 = \frac{1}{2}gpt^2$, from which, since g is a constant and p is the same for all planes of the same height, we have t proportional to s.

We may also prove the following theorem: The times of descent are equal down inclined planes whose lengths and inclinations are determined by the chords of a vertical circle drawn from its lowest point; or, otherwise stated, the time of descent down any chord of a vertical circle drawn from a point on the circle to its lowest point is the same as the time of free fall through the diameter of the circle. For, from the formula $s^2 = \frac{1}{2}gpt^2$, which was obtained in the last paragraph, the time t is the same for all inclined planes for which $\frac{s^2}{p}$ is the same; and $\frac{s^2}{p}$ is the same for all the chords of a vertical circle which are drawn to the lowest point.

To show this, consider a vertical circle (Fig. 53), and let NP be any chord drawn to the lowest point P of the circle. Upon the vertical diameter OP drop

from N the perpendicular NM. Now the triangles OPN and MPN are similar; whence, if we denote the diameter OP by d, the chord NP by s, and the length MP, the height of the inclined plane NP, by p, we have $d : s = s : p$, or $d = \frac{s^2}{p}$. Therefore $\frac{s^2}{p}$ is constant and equal to the diameter of the circle. Substituting for $\frac{s^2}{p}$ in the formula $s^2 = \frac{1}{2} g p t^2$ we obtain $t^2 = \frac{2d}{g}$ as the formula giving the time of descent down any chord drawn as described. The time t is the time of free fall through the diameter of the circle, and is the same for all the planes which conform to the given conditions.

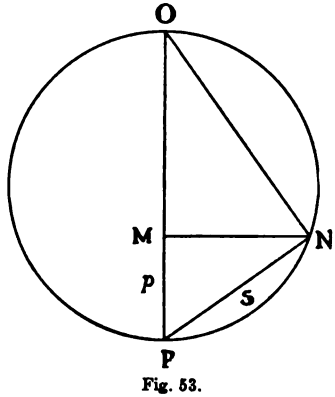


Fig. 53.

We may also prove the following theorem: The velocities acquired in the descent of a body down different inclined planes are equal if the heights of the planes are equal. For, from the formula $v^2 = 2as$ (Equations 19) and the relation $as = gp$, we obtain $v^2 = 2gp$, which shows that the acquired velocity depends only on the height of the plane and not on its length, and is therefore the same for all planes of equal height. This last relation holds even if the path of the moving body is curved, as, for example, if it is the arc of a circle; for, if not, the velocity acquired will either be greater or less than the velocity with which the body must start up an inclined plane in order to bring it to the level from which it started to fall. If the velocity thus acquired is greater, the moving body under the action of gravity alone will rise to a higher level than that from which it starts, and this conclusion is so contrary to our experience that we consider the hypothesis upon which it is based to be erroneous. If the velocity acquired is less, it is only necessary to reverse all the motions in order to reach the same erroneous conclusion from this supposition. We conclude, therefore, that when a body falls through any path, the velocity which it acquires in falling depends only on the perpendicular distance through which it falls and not on the length or shape of the path. Galileo illustrated this theorem by releasing a pendulum from a certain level, and stopping its cord, when it was passing through its middle position, by pins set at different heights (Fig. 54). The

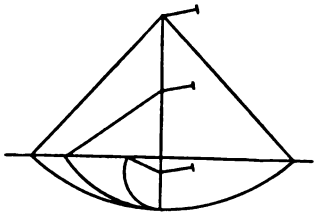


Fig. 54.

effective length of the pendulum was thus changed in the second half of its path, but, however the cord was stopped, the bob rose always to the level from which it started.

57. Composition of Velocities and Accelerations. — Galileo perceived or assumed that when a body is in motion and is then acted on by a force, the motion which it will subsequently have will be obtained by superposing on its original motion the motion given to it by the force. That is, the motion of the body is conceived of as consisting of two motions, which exist in the body independently of each other.

Let us suppose that a point is at one instant in a certain position and at a later instant in another position. The straight line which joins the first position to the second is called the *displacement* of the point. This displacement manifestly has a determinate direction as well as a magnitude, and it has no other characteristics. It is fully determined when the direction and magnitude are given. A quantity which conforms to these conditions is called a *vector*.

If a point undergoes two successive displacements, the final displacement which is the result of the two is obtained by drawing a line from the position first occupied by the displaced point to the position last occupied by it. This line is called the *resultant* displacement. It is manifestly the diagonal of the parallelogram of which the two displacements of the point are sides. What is true of displacements is also true of any similar vectors. The resultant of two or more vectors may be found, or a vector may be resolved into components, by the parallelogram law, or by those developments of it which have been given in the rules for compounding and resolving forces (§§ 41, 42).

Velocities, measured by displacements occurring in equal intervals of time, are vectors, and accelerations, measured by changes in velocity occurring in equal intervals of time, are also vectors of another sort. Velocities and accelerations may be compounded and resolved by the general rules which hold for the composition and resolution of vectors.

58. Vectors. — For such use as we shall make of them, vectors are generally represented in a diagram by finite lines drawn in definite directions, and in symbols by the letters in order designating the initial and final points of the lines with a bar placed over them. Thus $\overline{AB}$ represents the vector which is represented in Fig. 55 by the line drawn from A to B. The vector

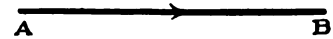


Fig. 55.

$\overline{BA}$ is represented in the diagram by the line drawn from B to A; and since this line is the reverse of the other, we have the sum or resultant of these two vectors

$$\overline{AB} + \overline{BA} = 0,$$

or

$$\overline{AB} = -\overline{BA};$$

so that a vector in one direction is equal to a vector of the same magnitude in the opposite direction with its sign changed.

By using the signs + and − with a special significance, we may write equations among vectors which can be handled by the ordinary operations of addition and subtraction.

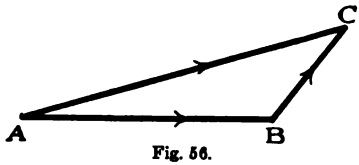


Fig. 56.

The expression

$$\overline{AC} = \overline{AB} + \overline{BC}$$

signifies that the vector $\overline{AC}$ (Fig. 56) is the resultant of the two vectors $\overline{AB}$ and $\overline{BC}$.

The expression

$$\overline{BC} = \overline{BA} + \overline{AC} = \overline{AC} - \overline{AB}$$

signifies that the vector $\overline{BC}$ is the difference between the vectors $\overline{AC}$ and $\overline{AB}$ in the sense that the vector $\overline{BC}$ added to the vector $\overline{AB}$, the subtrahend, will equal the vector $\overline{AC}$, the minuend.

The sign of equality signifies that the resultant vectors given by the operations indicated on either side of the equation are equal.

59. Projectiles. — Galileo applied the principles of the composition of motions to the study of the motion of projectiles.

When a body is thrown off in any direction as a projectile, its motion may be considered as made up of two independent motions, the motion originally given it, which is in the original direction of projection, and the motion imparted to it by the action of its own weight upon it, which is directed vertically downward.

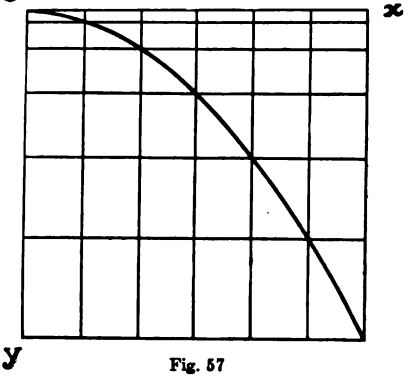


Fig. 57

We shall first consider the simple case of horizontal projection. Suppose a body to be projected horizontally (Fig. 57), from a point taken as origin, along the x-axis with the velocity u. At the end of the time t its displacement in the direction of the x-axis will be

$x = ut$. From the instant of projection it begins to fall toward the earth, and if we consider the y-axis to be directed vertically downward, the distance which it will fall in the time t in the direction of the y-axis is given by $y = \frac{1}{2}gt^2$. On the supposition which we have made, that the two motions can exist together in the body without mutual interference, we may obtain simultaneous values of x and y by eliminating t between these equations. The equation obtained is $y = \frac{g}{2u^2}x^2$ and represents the path of the body. It is the equation of a parabola with its vertex at the origin.

Since any projectile thrown obliquely upwards will reach a highest point of its path, and will at that point be moving horizontally, with a velocity equal to the horizontal component of the original velocity of projection, the path which it will subsequently traverse will be a parabola similar to the one already determined. Because of the symmetry of the conditions, the path of the projectile before it reaches its highest point will be the other branch of the same parabola.

To investigate the movement of a projectile in this more general case, we suppose it to be projected from the origin obliquely upwards (Fig. 58) with a velocity V, whose horizontal component along the x-axis is u and whose vertical component along the y-axis, directed upward, is v. Then in the time t the distance traversed along the x-axis is $x = ut$. The distance which would be traversed along the y-axis in the same time, if it were not for the change of velocity in that direction due to the weight of the body, is vt. Owing to the action of the weight of the body, the distance actually traversed in the direction of the y-axis is less than this, by the distance through which the body will fall in the time t. It is therefore given by $y = vt - \frac{1}{2}gt^2$. Eliminating t between these equations we obtain the equation $y = \frac{v}{u}x - \frac{g}{2u^2}x^2$ as the equation of the path of the body. It is the equation of a parabola which passes through the origin.

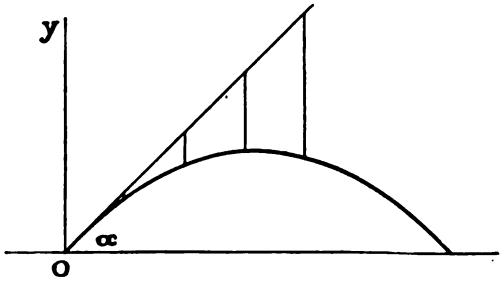


Fig. 58.

The *time of flight* is the time taken by the projectile, after leaving the origin, to fall to the same level again. It may be determined by setting $y = 0$ in the equation $y = vt - \frac{1}{2}gt^2$, connecting y with t, and determining the value of t obtained on this condition. We obtain $t' = \frac{2v}{g}$.

The *range* of the projectile is the distance between the origin and the point at which the projectile again meets the x-axis. It is therefore obtained by setting $y = 0$ in the equation $y = \frac{v}{u}x - \frac{g}{2u^2}x^2$, connecting y with x, and determining the value of x obtained on this condition. The range thus obtained is $x' = \frac{2uv}{g}$, and is the same, whatever the values of u and v may be, when they are such that their product is the same. If the initial velocity V is fixed, as well as the range, there are two directions of projection for which the product uv has the same value. These two directions are equally inclined to the line which bisects the right angle between the two axes. This may be shown most simply by representing u and v in terms of the initial angle of inclination α of the projection. We have $u = V\cos\alpha$, $v = V\sin\alpha$, and $2uv = 2V^2\sin\alpha\cos\alpha = V^2\sin 2\alpha$. Now $2uv$ is a constant, because the range is fixed, and V is also a constant by hypothesis, so that the only values that α can have, less than a right angle, are those which make $\sin 2\alpha$ a particular constant; and these are α and $\frac{\pi}{2} - \alpha$; for $\sin 2\alpha = \sin 2\left(\frac{\pi}{2} - \alpha\right)$.

For a given value of V, the inclination of projection for which the greatest range is obtained is the inclination of the bisector of the right angle between the two axes. For, the range is given by $x' = \frac{2uv}{g} = \frac{V^2\sin 2\alpha}{g}$; and for a fixed value of V this is a maximum when $\sin 2\alpha = 1$, or when $\alpha = 45°$.

60. The Pendulum. — An arrangement which, for practical reasons, attracted special attention in the early days of the study of mechanics is the pendulum. In the form in which it was first studied, the pendulum consists of a small heavy body, called the pendulum bob, swinging at the end of a long light thread firmly fastened at the top. Galileo concluded from observations that the time required for a pendulum to execute one swing is the same, whatever the extent of the swing may be. This conclusion is not strictly correct. Such observations as Galileo could make were not sufficiently accurate to detect its falsity, but the fact is that the time of swing of a pendulum is greater when the extent of the swing is greater. So long as the extent of the swing is kept within certain limits, so that the largest angle made by the suspending thread with the vertical is not greater than 10°, the times of swing are practically independent of their extent.

Galileo recognized the value of the pendulum as a means of measuring small intervals of time. Doubtless it was this use to which the pendulum could be put which directed special attention to it.

Huygens (1657) applied the pendulum to the regulation of clocks, and constructed clocks which are in all essential particulars

the same as those of the present day. He was led by his use of the pendulum to investigate its properties.

Huygens found that the property of isochronous oscillation which is possessed approximately by the ordinary pendulum swinging in a circular arc, is possessed exactly by a pendulum which is so adjusted as to swing in a cycloidal arc. For this reason the cycloid is called the tautochrone.

The problem presented by the pendulum is to find the way in which the time of oscillation depends upon the length of the pendulum.

61. Time of Oscillation of the Pendulum. — The problem of finding the time of oscillation of a pendulum was first solved by Huygens for the pendulum swinging in a cycloidal arc, and by inference from this case, for a pendulum swinging in a circular arc of small extent. We shall not attempt to follow the method of Huygens, but shall consider the motion of the ordinary pendulum, swinging in a circular arc, more directly.

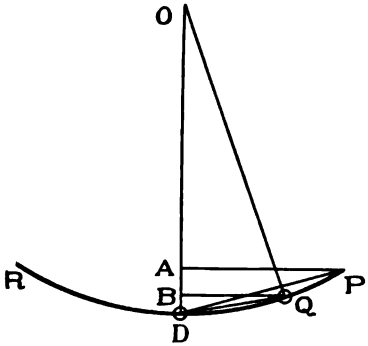


Fig. 59.

Suppose the pendulum suspended from the point O (Fig. 59) and swinging in the arc PDR. It will pass through any point Q with a velocity which we may obtain by the use of Galileo's theorem (§ 56) that the velocity of a body which falls through a vertical height s, whatever be the path over which it travels, is given by $v^2 = 2gs$.

If we represent by h the height AD of the highest level above D reached by the bob, when it is at the extremity of its arc at P, and by h_1 the height BD of the level of the point Q above D, we have by this theorem $v^2 = 2g\,(h - h_1)$.

Representing the chord DP by a, and the radius OD by r, we obtain from a well-known proposition in geometry,

$$h = \frac{a^2}{2r};$$

and representing the chord DQ by x, we obtain similarly

$$h_1 = \frac{x^2}{2r};$$

and hence

$$v^2 = \frac{g}{r}(a^2 - x^2). \qquad (21)$$

It is not possible for us, by elementary methods, to obtain the time of oscillation of a pendulum swinging in any arc, but by the help of this relation we may obtain the time of oscillation in an infinitesimal arc. Such an arc will differ infinitely little from a straight horizontal line, and we may represent it by a straight line in a diagram (Fig. 60).

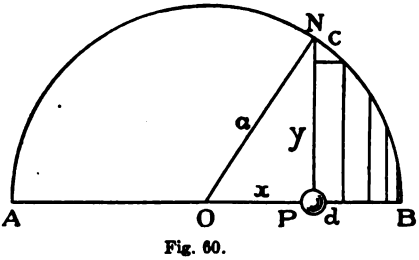


Fig. 60.

The half length of this line is a and the variable distance of the pendulum bob P from its center is x. If we conceive the line AB cut up into elements, or lengths so small that the velocity in any one of them may be considered constant, we shall have the small time Δt taken by the pendulum to traverse the element of length d, at the distance x from the center, given by $\Delta t = \frac{d}{v}$, in which the value of v is that given by Equation 21 for the value of x. The time of oscillation will be the sum of all the small time intervals in which the elements of the line AB are traversed. To find this sum we must make such a division of that line that the summation of all terms like $\frac{d}{v}$ is possible.

This is done by making use of the following construction: On AB as a diameter we construct a semicircle, and divide it into a large number of very small equal arcs of length c. These arcs are so small that any one of them may be considered a straight line. From the ends of these arcs we drop perpendiculars to the diameter AB and take the intercepts between these perpendiculars, or the projections of the various elements of arc, as the elements d of the line AB. The triangle of which c is the hypotenuse and d the base is similar to the triangle of which a is the hypotenuse and y the base; so that we have

$$d = \frac{c}{a} y.$$

From Equation 21 we have

$$v = \sqrt{\frac{g}{r}} \sqrt{a^2 - x^2} = \sqrt{\frac{g}{r}}\, y.$$

With these values of d and v we obtain

$$\Delta t = \frac{c}{a} \sqrt{\frac{r}{g}}.$$

Since c is a constant length, the time of description of each one of the elements d, of variable length, is the same.

To get the time of oscillation t we take the sum of all such terms for the whole semicircle. The factor $\frac{1}{a}\sqrt{\frac{r}{g}}$ appears in all the terms, and the sum of these terms therefore equals the sum of all the elements like c multiplied by $\frac{1}{a}\sqrt{\frac{r}{g}}$.

But the sum of all the elements like c is the arc of the semicircle, or is equal to πa; so that we obtain finally

$$t = \pi\sqrt{\frac{r}{g}}. \qquad (22)$$

In the use and study of the pendulum we call the time t in which the pendulum bob describes its path once, the *period* of the pendulum, or its *time of oscillation*. The radius r is called the *length* of the pendulum. The length a, the half-length of the path described by the bob, is called the *amplitude* of the oscillation. The distance x of the bob from the center of its path is called its *displacement*.

From the formula for the time of oscillation of a pendulum it appears that, since for any one place on the earth's surface the value of g is constant, the periods of different pendulums are proportional to the square roots of their lengths. At different places on the earth's surface the values of g are slightly different. In different places the times of oscillation of the same pendulum are inversely as the square roots of the value of g. Such differences in the time of oscillation of a pendulum, due to the different values of g at different places, were first observed by Richer when he transported a carefully rated clock from Paris to Cayenne (1671–1673).

62. The Physical Pendulum. — Real pendulums, especially those used in the regulation of clocks, are not made to conform to the definition of a pendulum which has been given. A real pendulum is a body of considerable mass, and is often irregularly shaped. If we consider each part of this pendulum as being itself the bob of a simple pendulum, it is plain that the times of oscillation of these different parts, if they were free from one another, would be very different. As they are bound together into one rigid body, some of them are forced to move faster and others are forced to move slower than they would if left to themselves. Manifestly there is some one of the parts which will swing in the same time

as it would if left to itself, or as if it were the bob of a simple pendulum, whose length is equal to its distance from the axis about which the pendulum swings. The problem of finding this length, which is called the length of the equivalent simple pendulum, was solved by Huygens. He employed, as the basis of his solution, the principle that the center of gravity of a system of bodies which are left to themselves can never rise to a higher level than that from which it started. This principle was questioned by Huygens's contemporaries, and its application to the solution of the pendulum problem anticipates certain conceptions which may better be brought out in another connection. We shall therefore postpone the discussion of Huygens's result until later.

63. Simple Harmonic Motion. — When we consider the pendulum swinging in any arc (Fig. 61), it is evident at once that the acceleration of the pendulum bob at any point Q of its arc is the same as the acceleration of a body on an inclined plane whose angle of inclination ϕ is the angle made with the horizontal by the tangent to the arc at Q. This angle ϕ is also the angle made with the vertical by the suspending cord. From the formula for the acceleration on the inclined plane, the acceleration is given by $f = -g \sin \phi$. The negative sign is introduced to indicate that the acceleration is in the negative direction when the angular displacement is positive, and in general is of opposite sign to the angular displacement.

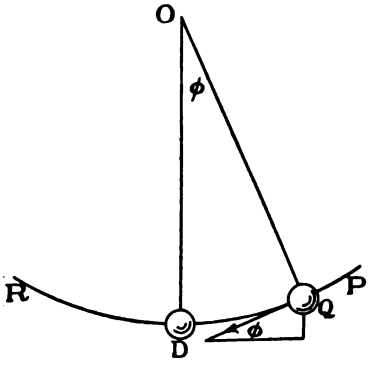


Fig. 61.

The acceleration is greatest at the ends of the swing; at the lowest point of the arc its value is 0, and its sign changes as the bob passes through that point, so that it is always directed so as to increase the velocity when the bob is moving toward the center and to diminish it when the bob is moving away from the center.

If the angle DOP is so small that the arc PDR may be considered a straight line, we may set the angle ϕ equal to its sine, and have

$$f = -g\phi.$$

Using x again for the chord of the arc DQ, we have $x = 2r \sin \frac{\phi}{2}$, and when the angle DOP is small $x = r\phi$.

Using this in the equation for the acceleration we obtain

$$f = -\frac{g}{r}x, \tag{23}$$

an equation expressing that the acceleration of the pendulum in an infinitesimal arc is proportional to its displacement, or distance from the center of the arc.

The motion of such a pendulum is a type of a kind of motion exhibited by very many bodies with greater or less precision, called *simple harmonic motion*. Its definition is that it is a motion in a straight line in which the acceleration is always proportional to the distance from a fixed point and directed toward that point. The amplitude and the displacement of the simple harmonic motion are defined as in the special case of pendulum motion, but the period T of the simple harmonic motion is the time of a double oscillation, and so the period of the pendulum is the half-period of the simple harmonic motion which it executes.

64. Simple Harmonic Motion, Continued. — We have defined a simple harmonic motion by the relation between the acceleration and the displacement, and from the study of the pendulum we have found an expression for the velocity in a simple harmonic motion in which the velocity is expressed as a function of the displacement. The motion, however, is not completely described until we have found the way in which the displacement, and so also the velocity and acceleration, depend upon the time.

We may do this by utilizing the construction by means of which we determined the period of the pendulum (§ 61). In Fig. 62 the moving point P is represented as having traversed the variable elements d lying between B and P. These elements are the projections of the equal elements c, and are described in equal times. If a point N moving out from B traverses the successive equal elements c in equal times it will move with constant speed, and the time in which it will traverse the arc BN will be to the time in which it will traverse the circumference of the circle as the length of the arc BN is to the circumference. If we represent the instant at which the point N starts from B by t_0, the instant at which it reaches N by t,

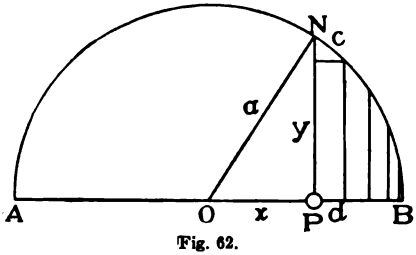


Fig. 62.

the period, or time of double oscillation, by T, and the angle subtended by the arc BN by ϕ, we may write this proportion in the form

$$t - t_0 : T = \phi a : 2\pi a,$$

from which we have

$$\phi = \frac{2\pi}{T}(t - t_0).$$

By the help of this angle we may write

$$x = a\cos\phi = a\cos\frac{2\pi}{T}(t - t_0). \tag{24a}$$

This formula gives us the displacement of the point P, which is the projection on the path AB of the point N moving uniformly in the circle, and which has the velocity and acceleration characteristic of the simple harmonic motion, in terms of the constants of the motion and of the time.

By using the special value for the period of the double oscillation of a pendulum, $T = 2\pi\sqrt{\frac{r}{g}}$, and substituting for $\frac{r}{g}$, and for $\sqrt{a^2 - x^2} = y = a\sin\phi$ in the formulas for the velocity and acceleration of the pendulum, we obtain

$$v = -\frac{2\pi}{T}a\sin\frac{2\pi}{T}(t - t_0), \tag{24b}$$

$$f = -\frac{4\pi^2}{T^2}a\cos\frac{2\pi}{T}(t - t_0). \tag{24c}$$

We insert the negative signs to indicate that the velocity is in the negative direction when $\sin\frac{2\pi}{T}(t - t_0)$ is positive, and that the acceleration is in the negative direction when $\cos\frac{2\pi}{T}(t - t_0)$ is positive.

These formulas contain only such quantities as are characteristic of simple harmonic motions in general, and fully describe such motions.

65. Uniform Motion in a Circle. — On account of the importance of the simple harmonic motion in the study of wave motions of all sorts and in electrical theory, we shall define it and determine its properties in another manner. To do this we shall need a theorem first announced by Huygens, descriptive of the acceleration of a point moving uniformly in the circumference of a circle.

Such a point has a variable acceleration, for although its speed is constant, the direction of its motion is continually changing. Huygens's theorem states that if a point is moving with constant speed v in a circle of radius r, its acceleration is directed toward the center of the circle, and is equal to $\frac{v^2}{r}$.

Let the moving point be at any instant at the point A (Fig. 63). In the short interval of time t succeeding, it will move to the point C on the circle, in an arc so short that it may be considered a straight line. This motion we conceive of as the resultant of two motions, one which it would have if it were not turned out of its original direction of motion at A, and which would carry it to B; the other a motion along the radius of the circle at A, and which would carry it to D. From the proportion among the sides of the similar triangles ACD and AEC, we have $AD : AC = AC : AE$. Setting $AD = s$, and remembering that when the arc AC is small the straight line AC is equal to it, and therefore to vt, we have from the proportion, $s = \frac{1}{2}\frac{v^2}{r}t^2$. Within the range set by our suppositions, therefore, and much more in the limit, as s and t approach zero as a limit, the distance traversed by the moving point toward the center of the circle is proportional to the square of the time. The motion in that direction is therefore uniformly accelerated (§ 54), and since for such a motion $s = \frac{1}{2}at^2$, we see that the acceleration a is given by the formula

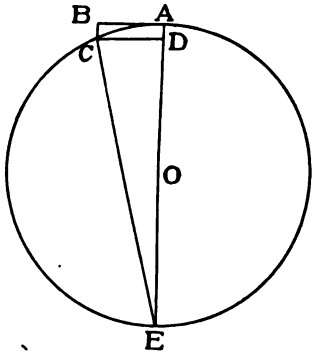


Fig. 63.

$$a = \frac{v^2}{r}. \qquad (25)$$

66. Simple Harmonic Motion, Resumed. — In the treatment of simple harmonic motion which we shall now consider, we define a simple harmonic motion as the motion in a limited straight line of the projection on that line of a point moving uniformly in a circle described upon it as diameter.

Suppose the point N (Fig. 64) to move uniformly in the circle on the diameter AB, with the constant speed V. It will describe

the circumference $2\pi a$ in the time T, the period; and we have $V = \frac{2\pi a}{T}$. In the same time the projection P of the point N will traverse the line AB twice. If we represent, as before, the instant at which the point N passes through the point B by t_0, the instant at which it arrives at N by t, the radius of the circle by a, the angle NOB by ϕ, and therefore the arc BN by ϕa, we have $\phi = \frac{2\pi}{T}(t - t_0)$. The distance x of the point P from the center, or the displacement, is then equal to

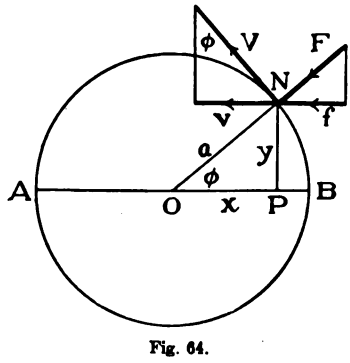


Fig. 64.

$$x = a \cos \frac{2\pi}{T}(t - t_0). \qquad (24a)$$

The velocity of the point P is the projection of the velocity $V = \frac{2\pi a}{T}$ of the point N, tangent to the circle, upon the line AB. This is manifestly equal to

$$v = -\frac{2\pi a}{T} \sin \frac{2\pi}{T}(t - t_0), \qquad (24b)$$

the negative sign being introduced because the velocity is directed in the negative direction so long as the sine of the angle is positive.

The acceleration of the point P is the projection of the acceleration of the point N, toward the center of the circle, upon the line AB. The acceleration of the point N, by the theorem of § 65, is given by $F = \frac{V^2}{a} = \frac{4\pi^2}{T^2} a$; and the acceleration of the point P is therefore

$$f = -\frac{4\pi^2}{T^2} a \cos \frac{2\pi}{T}(t - t_0), \qquad (24c)$$

the negative sign being introduced because the acceleration is directed in the negative direction when the cosine of the angle is positive, or when the displacement is positive.

The acceleration of the point P is therefore proportional to the displacement, and the motion thus defined possesses the characteristics of the simple harmonic motion as previously defined.

The ratio $\frac{2\pi}{T}$ is the angular velocity (§ 96) of the point N in the

circle. If we represent it by ω, and write ϵ for the constant ωt_0, the formulas representing the motion become

$$\left.\begin{aligned} x &= a\cos(\omega t - \epsilon), \\ v &= -\omega a \sin(\omega t - \epsilon), \\ f &= -\omega^2 a \cos(\omega t - \epsilon) = -\omega^2 x. \end{aligned}\right\} \tag{26}$$

The angle $\omega t - \epsilon$, which determines the motion, is called the *phase*, the angle ϵ, which depends upon the instant at which the moving point is at its greatest elongation, or is at the end of its path, is called the *epoch*, of the simple harmonic motion.

From the last of Equations 26, we have $\omega = \sqrt{-\frac{f}{x}}$, and since $\omega = \frac{2\pi}{T}$, we have for T the very convenient formula $T = 2\pi\sqrt{-\frac{x}{f}}$. In studying the motions of bodies or parts of bodies, we are often able to show that the acceleration is proportional to the displacement. In that case the motion is simple harmonic, and the formula just obtained enables us to find the period of the motion at once.

EXAMPLES, V

1. **Huygens's Problem:** — *To determine the time of fall down the arc of a vertical cycloid.*

Let B represent any point on the arc of the cycloid (Fig. 65), at which a body is released, to move under the action of its own weight down the arc to the lowest point O. Let d represent the length of a small element of the arc at any point C in the path. The time taken by the moving body to traverse the element d with the velocity v which it has at the point C will be

$$\Delta t = \frac{d}{v}.$$

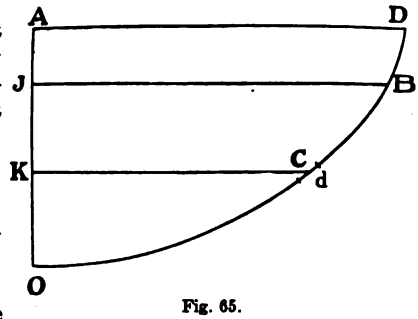


Fig. 65.

We get v from the formula (Equation 19)

$$v^2 = 2g(h - h_1),$$

in which h and h_1 are the heights of the points B and C above the lowest point O and $h - h_1$ is therefore the vertical distance between the levels of B and C.

Now from the geometry of the cycloid, the height OJ of any point B on it is connected with the distance OB measured from the vertex to B along the arc by the formula

$$4l\,OJ = OB^2,$$

in which l is the diameter OA of the generating circle, or the height of the highest point D above the vertex. Using this relation to replace h and h_1 by the corresponding arcs a and x, we have

$$v^2 = \frac{g}{2l}(a^2 - x^2),$$

and hence

$$\Delta t = \sqrt{\frac{2l}{g}} \frac{d}{\sqrt{a^2 - x^2}}.$$

The ratio $\frac{d}{\sqrt{a^2 - x^2}}$ is similar to the one of the same form which appeared in the discussion of the pendulum (§ 61), and may be interpreted in a similar manner, as the ratio of an infinitesimal element c of the arc of a semicircle to its radius a. Replacing it therefore by $\frac{c}{a}$, and adding all the similar time intervals in which the different elements of the arc BO are traversed, we have $\Sigma c = \frac{\pi a}{2}$, and $t = \frac{\pi}{2}\sqrt{\frac{2l}{g}}$.

This is the time of fall down the arc of the cycloid to the lowest point. No restriction has been made in the demonstration to an infinitesimal portion of the arc; and since the time here found is independent of the length of the arc, it will be the same from whatever point on the arc the body is released.

Since the motion of the body in the other branch of the cycloid, after it has passed the lowest point, is the counterpart of the motion in the arc BO, the time of a complete oscillation in the cycloidal arc will be

$$T = \pi\sqrt{\frac{2l}{g}},$$

and will be independent of the amplitude of the oscillation.

The radius of the circle which touches the cycloid at its vertex is $2l$. Calling this r we have

$$T = \pi\sqrt{\frac{r}{g}},$$

the time of oscillation of a body moving in that infinitesimal portion of the circular arc of radius r which is practically coincident with the arc of the cycloid.

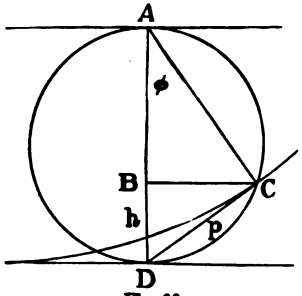


Fig. 66.

2. *To find the acceleration of a body moving in a cycloidal arc.*

The acceleration at any point C on the cycloid (Fig. 66) is $g \sin \phi$, where ϕ is the inclination of the tangent at C to the horizontal. If a circle of diameter l (the diameter of the generating circle) is drawn touching the horizontal line drawn through the lowest point of the cycloid at the point where it is intersected by the tangent from C, it will pass through C, and the line AC will be perpendicular to the tangent of the cycloid at C, so that the angle $DAC = \phi$. Representing the chord DC by p, we have $\sin \phi = \frac{p}{l}$. Representing BD by h and the arc of the cycloid from C to its lowest point by s we have $h = \frac{p^2}{l}$, and from the property of the cycloid already used $h = \frac{s^2}{4l}$, so that $p = \frac{s}{2}$. Using this value of p in the expression for the acceleration we have

$$g \sin \phi = g \frac{p}{l} = \frac{g}{2l} s.$$

The acceleration in the arc of the cycloid is proportional to the distance along the arc from its lowest point. The motion in a cycloid is therefore similar to a simple harmonic motion in its essential features.

3. **Circular or Angular Harmonic Motion.** *To find the motion of a body which moves in the arc of a circle with an acceleration along the arc proportional to the length of the arc, and directed toward the point from which the arc is measured.*

Let x represent the length of the arc. Then the acceleration f will be represented by $f = -\omega^2 x$. This is the fundamental relation for harmonic motion.

If we reckon time from the instant at which the body is at the extreme end of the arc, so that $t_0 = 0$, the displacement measured along the arc can be represented by $x = a \cos \frac{2\pi t}{T}$ (Equation 24a), the velocity in the arc by $v = -\frac{2\pi}{T} a \sin \frac{2\pi t}{T}$ (Equation 24b), the acceleration along the arc by $f = -\frac{4\pi^2}{T^2} a \cos \frac{2\pi t}{T}$ (Equation 24c). The amplitude a is the length of the extreme arc described.

If we divide these equations by r, the radius of the arc, and write $\phi = \frac{x}{r}$, $\omega = \frac{v}{r}$, $\alpha = \frac{f}{r}$, $A = \frac{a}{r}$, we shall obtain $\phi = A \cos \frac{2\pi t}{T}$, $\omega = -\frac{2\pi}{T} A \sin \frac{2\pi t}{T}$, $\alpha = -\frac{4\pi^2}{T^2} A \cos \frac{2\pi t}{T}$. These quantities are the angular displacement (§ 95), angular velocity (§ 96), and angular acceleration (§ 97), of the radius drawn to the body. The angular acceleration is proportional to the angular displacement, and the motion has the general characteristics of harmonic motion. It is called angular harmonic motion.

4. **Forced Vibrations.** A body is so situated that, when it is slightly displaced, a force acts upon it which produces an acceleration proportional to its displacement and oppositely directed to it. We write this relation in the form $f = -\omega^2 x$. The motion which results is simple harmonic, of the period $T = \frac{2\pi}{\omega}$ (§ 66). It is called a free vibration. Another force is applied to the body to move it in the same path, which changes according to the simple harmonic law, with another period S. We shall consider the resultant motion.

The applied force will produce an acceleration which is proportional to it, and may be expressed by $e \cos \frac{2\pi t}{S}$. The resultant acceleration, arising from the addition of the acceleration due to the conditions of the body and that due to the applied force, is $f = -\omega^2 x + e \cos \frac{2\pi t}{S}$.

We cannot solve this equation directly, by such methods as we can use, but we can obtain a solution for it by inspection that will describe a possible motion. To do this we set $x = z \cos \frac{2\pi t}{S}$, that is, we assume that the displacement follows the simple harmonic law, with the period of the applied force, but with an amplitude z unknown, and to be found. Substituting for $\cos \frac{2\pi t}{S}$ in the equation of

motion we get $f = -\left(\omega^2 - \frac{e}{z}\right)x$. The resultant motion is simple harmonic, with the period $S = \frac{2\pi}{\sqrt{\omega^2 - \frac{e}{z}}}$. Substituting for ω^2 its value $\frac{4\pi^2}{T^2}$, we obtain for the displacement x the equation

$$x = z \cos \frac{2\pi t}{S} = \frac{e}{4\pi^2} \frac{S^2 T^2}{S^2 - T^2} \cos \frac{2\pi t}{S}.$$

The resultant amplitude depends upon the relative magnitudes of S and T. If the period S of the impressed force is greater than the period T of the free oscillation, the amplitude is positive, and the phase of the resultant motion is the same as that of the impressed force. If S is less than T, the amplitude is negative, or the phase of the resultant motion is opposite to that of the impressed force.

As S approaches T in value, the amplitude increases without limit. In practice, an infinite amplitude is not attained, because of friction and other similar disturbing causes, but with a close agreement between the two periods the amplitude may become very great.

Such a motion is called a forced vibration. It may be exhibited by suspending a pendulum with a light bob from the heavy bob of another pendulum. If the period of the heavy pendulum is greater than that of the light one, the two pendulums will swing in the same sense. If it is less, they will swing in opposite senses. As the two periods are adjusted to be nearly equal, the vibrations of the light pendulum become very great.

Mass and Momentum

67. Newton's Laws of Motion. — We have now attained certain fundamental conceptions respecting the motion of a body when subjected to the action of a force. One of these conceptions is, that motion, as well as rest, is a natural state of a body, and that a body will persist in its motion without change except in so far as it is acted on by a force. This conception may be called the *principle of inertia.*

The way in which it was reached by Galileo is of special interest, although his process of thought cannot be called a proof of the principle. Galileo conceived a body to move down an inclined plane and at the bottom of the plane, by a change in direction which did not involve a change in the body's speed, to begin an ascent on a second inclined plane. In accordance with the propositions in § 56, the body will move along the second inclined plane over a length which will bring it to the same level as that from which it started to fall, and the time of ascent through this length will be proportional to the length. Now, as the inclination of the second plane is made less and less, the time of ascent will become longer and longer and the retardation of the body, or its change of velocity in a unit of time, will be less and less. In the limiting case, in which the inclination of the plane is zero, the time of ascent becomes infinite and there is no retardation. In this special case, therefore, in which the moving body is not acted on by any force, its motion will be uniform and indefinitely continued.

Another of these fundamental conceptions is that the result of the application of a constant force to a body is the production of a constant acceleration. The constant force considered by Galileo was the weight of the body or a component of its weight. Galileo's experiments indicated that the accelerations produced in the same body by different components of its weight, that is, the accelerations of the body when moving down planes of different inclinations, are proportional to those components. That is, they indicated for forces arising from a particular cause that the forces and the accelerations which they produce in a body are proportional. This relation, however, was not stated, or even assumed, as a general one involving the action of all kinds of forces, perhaps merely because the forces considered by Galileo were all of one kind.

In his *Philosophiae Naturalis Principia Mathematica*, published in 1687, Newton collected the scattered and imperfectly formulated principles of his predecessors, and adding to them an additional principle, stated them in three *laws of motion*, which furnish a complete basis for the science of mechanics.

Newton began by defining the quantity of matter in a body as a magnitude measured by the product of the density and the volume of the body. He named this quantity the *mass* of the body. He did not intend this definition to be final, nor did he think that it furnished an adequate means of determining different masses. He merely wished to call attention to the fact that a certain measurable magnitude exists for each body, which will depend on the nature of the matter contained in each unit of volume and on the volume occupied by the body.

Newton next defined a quantity to which we now give the name *momentum*. The momentum of a body is the product of its mass and its velocity. We express it in symbols, using m to represent the mass of the body measured in some appropriate unit, by the product mv.

After distinguishing different ways in which forces may be considered and defined, Newton presented the laws of motion as follows:

Law I. Every body continues in its state of rest or of uniform motion in a straight line, except in so far as it is compelled by an external force to change that state.

Law II. Change of momentum is proportional to the external force applied and takes place in the direction of the applied force.

Law III. To every action there is an equal and contrary reac-

tion; or, the mutual actions of two bodies are always equal and oppositely directed.

These laws are physical axioms or postulates. They are not self-evident truths; they can hardly be said to be immediately demonstrable by experiment; they are rather generalizations which are so consistent with the experience of every one who fairly considers them that their correctness is readily admitted as probable. The final test of their correctness is in the agreement of all the consequences which flow from them with the results of observation.

68. The First Law. — The first law simply states the principle of inertia as it was understood by Galileo. It is in a sense contained in the second law, for it is an evident implication of the second law that, if no external force is applied to a body, the motion of the body will not change.

Since the case of a body at rest is obviously included under the more general one of a body moving with a constant velocity, we may state the first law in harmony with our previous definition of constant velocity by saying:

The velocity of a body is constant except while it is acted on by a force arising from the action of other bodies.

69. The Second Law. — In the second law Newton meant by "change of momentum" the change of momentum per unit of time, or what we now call the rate of change of momentum. Since the mass of a body is not changed by its motion or by the action of a force on it, a change of momentum involves only a change of velocity; and the rate of change of momentum of a body is the product of its mass and its acceleration.

The second law may therefore be expressed in the formula

$$F \propto ma.$$

Now the unit of mass has been left undefined, the symbol m being simply used to represent mass measured in some unit. In the proportion $F \propto ma$, F and a are both familiar quantities, measured in units which we have already adopted; and so if we consider m as the factor of proportion between F and a and write

$$F = ma, \tag{27}$$

we have a fundamental equation expressing the significance of the second law, and at the same time defining the unit of mass.

The second law recognizes the independence of forces. When more than one force acts on a body, each force produces its own independent effect, by imparting an acceleration to the body in the

direction in which it is acting, which is the same as it would be if no other forces were acting on the body.

We may paraphrase the second law in the following statements:

The acceleration of a body caused by a force is proportional to the force and in its direction.

The accelerations caused in different bodies by equal forces are inversely as their masses.

Each force acting on a body causes an acceleration such as it would cause if other forces were not acting.

70. The Parallelogram of Forces. — From the second law we may deduce the parallelogram of forces. When two forces act on a body at once, they will give to the body accelerations in the directions of the forces and proportional to them. These accelerations are vectors and may be added by the parallelogram law, and since the forces which produce them are proportional to them, and in their directions, the forces are also vectors and may be added by the same law. This so-called proof of the parallelogram of forces is not logically more complete or convincing than the direct experimental proof which has already been given, for it rests ultimately on the experimental evidence for the truth of the second law.

71. Relation of Mass to Weight. — We may illustrate the statements of the second law with the Atwood's machine (§ 53). With the main weights made up, as they generally are, by assembling a number of smaller weights of equal volume and of the same material, and with the overweights adjusted to them, we may determine the mass of the moving system by simply counting the number of small equal masses which make it up, without as yet being able to measure mass in terms of a final or absolute unit. When we do so and try the effect of different overweights on the same mass, we find that the accelerations produced are proportional to the overweights. When we try the effect of the same overweight on different masses we find that the accelerations are inversely as the masses.

And further, if we construct equal weights of different materials and use them in turn as the masses of the system, we find that the same overweight will produce the same acceleration in each case; and in general, different weights of different materials used with the same or with different overweights will behave with respect to the accelerations which they will have as if their masses were proportional to their weights.

The fact that bodies of different weights fall toward the ground with the same acceleration confirms this conclusion. For if, in the

equation $F = mg$, which applies to each of these different bodies, the acceleration g has a common value and the force F is equal to the weight of the body in each case, the mass m must be proportional to the weight of the body in each case.

The final proof of the relation indicated by these experiments is obtained from experiments with the pendulum, as will be explained later.

72. Measurement of Mass. — The relation between force and mass which is expressed in the formula $F = ma$, is, from the point of view which we have taken, a relation which defines and measures mass. All that was uncertain and illogical in Newton's preliminary definition has been replaced by a precise and intelligible connection of the quantity, mass, to be measured, with other quantities which we already know how to measure. Any mass may be measured by observing the acceleration produced in it by the action on it of a known force. A standard force may be selected, and then a standard or unit mass determined, and all other masses measured by the use of it. In fact, however, we proceed differently in establishing the unit of mass. By taking advantage of the relation stated in the third law we can compare masses directly by means of their accelerations, without knowing the magnitude of the forces which act on them. Having thus measured masses in terms of some one mass arbitrarily chosen as standard, we use the formula $F = ma$ as the relation which enables us to measure forces.

73. The Third Law. — Up to the time of Newton no strict inquiry had been made as to the source or origin of force. The physical universe was supposed to consist of matter and of forces, of which the origin was unknown and did not need to be known. The subject of study was the motion of a body under the action of given forces. Newton perceived that a force which acts on a body always arises from the action of some other body, and further that this other body is also acted on by a force arising from the first body. Thus all action on a body arises from an interaction between it and other bodies. The third law states, in part, that the elements which constitute this interaction, that is, the forces which act on the two bodies, are equal and oppositely directed.

Newton used the words "action" and "reaction" in a double sense. Primarily they mean force, and interpreted as force the law holds true in every case. But secondarily they mean also the result of the action of a force in change of momentum. When bodies exert force on each other, whether they move or not, the

forces exerted are equal and opposite. When the bodies move under the action of these forces, their changes of momentum are equal and opposite.

We may express the meaning of the third law by saying:

All forces arise from the interaction of bodies. When two bodies interact, each is acted on by a force, and the two forces composing the interaction are equal and opposite in direction. If the two bodies experience changes of momentum these are also equal and opposite in direction.

74. Comparison of Masses. — We are now in position to compare masses with each other without the use of any known or measured force. We need only allow two masses to interact with each other, either directly by collision, or by the intervention of some other body, and observe the changes of their momenta. According to the third law their changes in momentum will be equal and of opposite sign. If we suppose the masses m_1 and m_2 to be moving in the same line, and represent by v_1 and v_1', v_2 and v_2', their velocities before and after their interaction, we may write

$$m_1 (v_1 - v_1') = m_2 (v_2' - v_2) \tag{28}$$

or

$$\frac{m_2}{m_1} = \frac{v_1 - v_1'}{v_2' - v_2}.$$

All other masses may be compared in this way with the mass m_1, and the values of these masses found in terms of m_1 taken as a standard or unit.

If the masses under observation are compared with the mass m_2, or with any other one of the set, the numerical values for them found in terms of the one taken as standard will be different in each case; but the ratios of the values of any two masses expressed in terms of different standards will be the same.

Because of the permanence of matter and of its unchangeableness with change of place, it is found desirable to adopt as one of the fundamental units, or arbitrary standards from which our absolute system of units is built up, a unit or standard mass.

75. Unit of Mass. — The body whose mass is taken as a standard or unit of mass is a certain piece of platinum, called the standard kilogram. The standard pound, or the mass of a certain other standard piece of platinum, is also often used as a unit of mass by the English-speaking nations (§ 5).

It has been shown by experiment that equal masses, whatever the material of which they are composed, have equal weights, and it is

therefore most convenient to construct sets of standard masses and their multiples and submultiples by weighing them. In physical work, the unit mass which is most commonly employed is the one-thousandth part of the standard kilogram. This mass is called the gram.

76. Units of Force. — In our preliminary choice of a unit of force for use in the study of statics, we selected the weight of the standard mass. This unit is not used as an absolute unit, and we shall now discard it and replace it by a unit defined in terms of the three fundamental units. To define the unit of force in an absolute system we recur to the formula (Equation 27) $F = ma$; for now having measured mass in terms of the unit of mass, and being able to measure acceleration in terms of the units of length and time, we are able to measure force also in these units; and the unit of force indicated by this formula is independent of circumstances of time or place, at least in so far as the unit of mass is so independent, and is an absolute unit.

The absolute unit of force is the force which will impart unit acceleration to the unit mass.

The unit of force in the c.g.s. system is that force which will impart to a gram one unit of acceleration, that is, it is that force which, acting uniformly for one second, will impart to a gram a velocity of one centimeter per second. This unit of force is called the *dyne*. The force which will impart to a pound in one second a velocity of one foot per second is also taken as a unit force. It is called the *poundal*.

In practice, especially in engineering practice and on a large scale, forces are often measured in terms of what is called the *gravitation unit*. This unit is the weight of a standard mass. This standard mass may be the gram, but is more often the kilogram, the pound or the ton. The numerical value of a force in terms of the gravitation unit is its ratio to the weight of the standard mass. Since the weight of a body is different at different places on the earth, the gravitation unit is not everywhere the same; but it differs so little in different places, that its variations may usually be neglected in practice.

The name of the gravitation unit is the name of the mass whose weight is taken as the unit. Thus, we speak of the weight of a kilogram, or simply of a kilogram, as a unit force.

The weight of a body in absolute units is equal to the mass of the body multiplied by the acceleration which will be imparted

to it by its weight, that is, $W = mg$; and hence the gravitation unit, that is, the weight of the unit of mass, is equal to g absolute units of force. The value of g at latitude 40° is about 980, when the centimeter and the second are the units of length and time. The weight of a gram is therefore about 980 dynes. When the foot and second are taken as the units of length and time the value of g is roughly equal to 32. The weight of a pound is therefore equal to 32 poundals.

77. Impulse. — It often happens that a body is acted on by a force applied for so short a time, and perhaps in so irregular a manner, that it is not possible to determine its values during the time of application, so as to calculate from them the effect upon the body. Such a force is applied by a blow of a hammer or by the explosion of powder in a gun.

To discuss this case, we consider the short time in which force is acting on the body as divided into a large number of intervals $\tau_1, \tau_2, \ldots \tau_n$, each one of which is so small that the force which acts on the body during it may be considered constant. Representing by $F_1, F_2, \ldots F_n$ the forces acting in the corresponding intervals, by v_0 the initial velocity of the body, by $v_1, v_2, \ldots v_n$ the velocity of the body at the end of each interval, and by m the mass of the body, the general formula connecting force and mass will then give

$$F_1 = m\frac{v_1 - v_0}{\tau_1},$$

$$F_2 = m\frac{v_2 - v_1}{\tau_2},$$

$$\cdots\cdots\cdots$$

$$F_n = m\frac{v_n - v_{n-1}}{\tau_n},$$

or

$$F_1\tau_1 = mv_1 - mv_0,$$

$$F_2\tau_2 = mv_2 - mv_1,$$

$$\cdots\cdots\cdots$$

$$F_n\tau_n = mv_n - mv_{n-1},$$

and by addition

$$\Sigma F\tau = mv_n - mv_0. \qquad (29)$$

The sum $\Sigma F\tau$, properly the time integral of the force, is called the *impulse*. It is equal to the change in momentum of the body.

It is customary to call a force applied to a body in the way which has been described an impulsive force. While it is not possible to observe it directly, the impulse in any case may be determined by

observing the change in momentum produced by it. We may then obtain some information about the force which has acted by remarking that if t represents the whole time during which force has acted on the body, so that

$$t = \tau_1 + \tau_2 + \ldots \tau_n,$$

we may obtain the average value F of the force from the formula

$$F = \frac{\Sigma F\tau}{t} = \frac{mv_n - mv_0}{t}.$$

The impulsive force in most cases obviously ranges from an initial value zero to a final value zero. The average force gives a lower limit to the actual force applied for at least a part of the time t.

The unit of impulse is the impulse of a unit force acting for unit time. In the c.g.s. system it is the impulse of a dyne acting for one second. No name has been given to it.

78. Centripetal Force. — In § 65 it was shown that a point moving with constant speed v in a circle of radius r has an acceleration directed toward the center of the circle equal to $a = \frac{v^2}{r}$. If a body of mass m moves similarly in the circle, it is acted on by a force equal to $\frac{mv^2}{r}$, directed toward the center of the circle. This force may arise from the action of a body placed at the center, from the tension in a cord or other similar body joining the moving body with the center, from the pressure of a circular wall, or in other ways. As it is always directed toward the center, it is called a *centripetal force.*

From the equality between a force and the product of the mass upon which it acts and the acceleration caused by it, it is evident that if we substitute for this product a force equal to it in magnitude, and opposite in direction to the acceleration, we shall have a system of two forces in equilibrium. Applying this conception to the case under consideration, the fictitious force which we thus introduce is equal to the centripetal force and is directed away from the center of the circle. It is equal to, and in the same direction as, the real force or reaction which is exerted upon a body at the center of the circle, when the centripetal force is due to the action of such a body. The observation of this reaction, which is a real force, has led to a confusion between it and the fictitious force which is equal to it, and it is often thought that the moving body is acted on by a force tending to carry it away from the center.

This supposititious force is called a *centrifugal force.* Strictly speaking, there is no such force as a centrifugal force, but the term is often a convenient one if it is understood to mean simply the fictitious force which may be substituted for the product of the mass of the body and its acceleration.

79. Motion in any Path. — When a body is moving in any path, its acceleration is generally variable. The acceleration may be resolved into two components, tangent and normal to the path. The tangential component a_t is equal to the rate at which the numerical value of the velocity changes in the path, so that if v and v' represent two values of the velocity taken at the beginning and end of an interval of time t so short that the rate of change of velocity during it is constant, we shall have

$$a_t = \frac{v' - v}{t}.$$

The normal component a_n is given by the formula of the last section; for an infinitesimal arc of any curve coincides with the corresponding arc of its osculating circle, so that if we represent by r the radius of that circle, or the radius of curvature of the path at any point, the normal acceleration at that point is given by $a_n = \frac{v^2}{r}$.

The general acceleration of a moving body is the resultant of these tangential and normal components. When the path is straight, the radius of curvature is infinite and there is no normal acceleration. When the velocity remains numerically constant the only acceleration possible is one of constant magnitude directed toward a fixed point, and the motion is uniform motion in a circle.

The force which acts on the body and causes this motion may similarly be resolved into tangential and normal components. These components will be equal to the mass of the body multiplied by the tangential and normal accelerations respectively.

EXAMPLES, VI

In dealing with the motions of bodies which are joined together in a system, the general method of determining them consists in obtaining, by inspection of the given conditions, the resultant force which acts on each body of the system, and setting this equal to the mass of the body multiplied by its acceleration. The equation thus formed is called the *equation of motion* of the body. If the resultant force is fully known, the acceleration of the body can be found at once from this equation.

It often happens that one or more of the component forces which make up a resultant force are the actions and reactions between the bodies of the system.

These actions and reactions, called more commonly, in this connection, the tensions between the bodies, are unknown, and must be eliminated between the equations of motion of the bodies to determine the motion. This can be done if sufficient conditions are given.

This general method may often be simplified in special cases.

1. **The Atwood's Machine.** In treating the Atwood's machine at this time we shall neglect the effect of the mass of the wheel upon the movement of the system. The masses moved are then M_1 and M_2, of which M_1 is the greater by the mass of the overweight. There is a tension T in the cord which acts upward on both masses alike. The common acceleration may be called a. Then the equations of motion are

$$T - M_2 g = M_2 a, \quad M_1 g - T = M_1 a,$$

from which we get $a = g\dfrac{M_1 - M_2}{M_1 + M_2}$, $T = g\,2\dfrac{M_1 M_2}{M_1 + M_2}$.

2. Representing the M_1 of Example 1, by $M_2 + m$, where m is the mass of the small overweight, we see by inspection of the formula for a that when the main masses are large the acceleration of the system is approximately proportional to the overweight.

3. *A smooth horizontal table is furnished with a pulley clamped at one end of it. A cord passes over the pulley and its ends are fastened to two masses, one of which, M_1, hangs down, while the other, M_2, rests upon the table. The system of masses is released, and it is desired to find the motion.*

Calling the tension in the cord T, and the acceleration a, the equations are

$$T = M_2 a, \quad M_1 g - T = M_1 a;$$

whence $$a = g\frac{M_1}{M_1 + M_2}, \quad T = g\frac{M_1 M_2}{M_1 + M_2}.$$

This may easily be seen otherwise, for evidently the moving force is $M_1 g$, and the mass moved is $M_1 + M_2$, from which the equation for a follows directly.

When the mass M_1 is small, the tension is approximately proportional to its weight.

4. *Find the motion of a similar system when the table is inclined at the angle α with the horizontal, and the pulley stands at the top of the slope.*

The equations are

$$T - M_2 g \sin\alpha = M_2 a, \quad M_1 g - T = M_1 a;$$

whence $$a = g\frac{M_1 - M_2 \sin\alpha}{M_1 + M_2}, \quad T = g\frac{M_1 M_2 (1 + \sin\alpha)}{M_1 + M_2}.$$

These equations reduce to those of Example 3, if the plane is horizontal, to those of Example 1, if the plane is vertical.

5. **The Conical Pendulum.** — A pendulum bob suspended by a fine cord that can swing with equal freedom in any azimuth may be drawn aside and projected perpendicularly to the vertical plane containing the cord, with such a particular velocity that its path will be a circle with its center on the vertical line under the point of support. The cord describes a right cone, and the pendulum thus employed is called a conical pendulum. The problem is to find the expression for its period, that is, the time in which it describes its circular path.

We represent by l the length of the cord, by ϕ the angle between the cord

and the vertical. The forces which act on the bob P (Fig. 67) are its weight $W = mg$ and the tension R in the cord. Since the bob moves in a circle of radius $r = l \sin \phi$, the force acting toward the center of the circle is $\frac{mv^2}{r}$ (§ 78). This must be equal to the resultant of the two forces R and W along the radius, or to $W \tan \phi = mg \tan \phi$. We have therefore for the velocity v the equation $v^2 = gr \tan \phi$. Now the bob describes a circle of circumference $2\pi r$ in its period T, so that $v = \frac{2\pi r}{T}$. With this value for v, and substituting l for $\frac{r}{\sin \phi}$, we get $T = 2\pi \sqrt{\frac{l \cos \phi}{g}}$. This is the time taken for a simple pendulum of the length $l \cos \phi$ to swing to-and-fro in its arc. This theorem was discovered by Huygens.

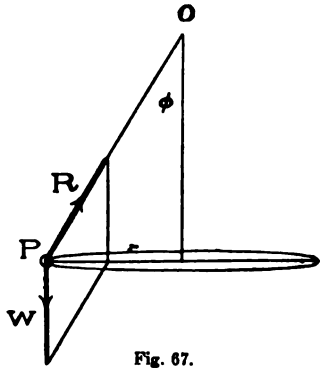


Fig. 67.

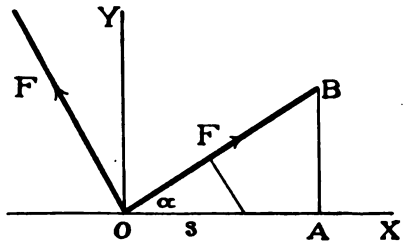


Fig. 68.

Work and Energy

80. Work. — The work of a force has been defined (§ 33) as the product of the force and of the component of the distance through which its point of application moves in the direction of the force. If the direction of the force (Fig. 68) and the direction in which its point of application moves are inclined to each other by the angle α, the work of the force is expressed by the formula $Fs \cos \alpha$, in which F represents the force and s the distance through which its point of application moves. This formula may also be interpreted by saying that the work of a force is measured by the product of the distance through which the point of application moves and the component of the force in the direction of motion. According to the value of the angle α, the work of the force may have any value between Fs and $-Fs$. When the force and the motion of its point of application are perpendicular to each other, the work of the force is zero. When the angle α is less than 90°, so that $\cos \alpha$ is positive, the work of the force is positive. In this case we say that work is done by the force. When the angle α is greater than 90°, so that $\cos \alpha$ is negative, the

work of the force is negative. In this case we may say either that negative work is done by the force or that work is done against the force.

81. Units of Work. — The absolute unit of work is the work done by unit force when its point of application moves through unit distance in the direction of the force. The work done by a dyne when its point of application moves in its direction through one centimeter is the unit of work in the c.g.s. system. It is called an *erg*. For many uses the erg is inconveniently small; and a larger unit, called the *joule*, equal to 10^7 ergs, is employed instead of it.

In practice it is often found convenient to use as a unit of work the work done by the gravitation unit of force when its point of application moves vertically through a certain distance chosen as unit distance. Thus the work done by a kilogram falling vertically through one meter is often taken as a unit of work. It is called the *kilogram-meter*. The *foot-pound*, or the work done by a pound falling vertically through one foot, is also a unit of work which is often employed by the English-speaking nations.

The rate of doing work is a quantity of great importance in the consideration of the value of machines. It is called the *power*, and is measured by the number of units of work done in unit of time. The absolute unit of power is the power equivalent to the doing of unit work in unit time. In the c.g.s. system it is the power equivalent to the doing of one erg of work in one second. The power equivalent to the doing of one joule of work in one second, named the *watt*, is often used in practice.

The *horse power*, or the power which does 33,000 foot-pounds of work in one minute, is a unit frequently used in engineering. It is equal to 746 watts.

82. Kinetic Energy. — When a constant force F acts on a free body and moves it from rest through any distance s, it does work equal to the product of the force and the distance, or to Fs. In this case, where there are no counteracting forces, the force is equal to the product of the mass m of the body and the acceleration a imparted to it, or $F = ma$. From § 54, Equation 20c, a body which enters upon a distance s with the velocity v_0 and moves through that distance with the constant acceleration a, will leave it with the velocity v given by the formula $v^2 - v_0^2 = 2as$. Multiplying both sides of this equation by m, we obtain $mv^2 - mv_0^2 = 2mas$, or

$$Fs = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2. \qquad (30)$$

This result expresses the relation between the work done by a force upon a free body and the velocity which the body will acquire.

The quantity expressed by $\frac{1}{2}mv^2$ is called the *kinetic energy* of the body. It is manifestly not a vector, that is, it does not depend in any way on direction, and it has no negative values. The result of the argument just given is expressed in the statement that the work done on a free body by a constant force is equal to the kinetic energy gained by the body.

We may extend the preceding demonstration to the case of variable forces and of a body in motion at the beginning of the path in which the work is measured. Represent by $s_1, s_2, \ldots s_n$, portions or elements of the path of the body, so short that the force acting on the body as it traverses each of them may be considered constant. Represent by $F_1, F_2, \ldots F_n$ the components of the forces which act in the corresponding elements of path. Represent by v_0 the initial velocity of the body along the path as it enters upon the first element, and by $v_1, v_2, \ldots v_n$ the velocity with which it leaves each element. Then from the relation of Equation 20c $v^2 - v_0^2 = 2as$. Writing out this equation for each element and multiplying by m, we have

$$2ma_1s_1 = mv_1^2 - mv_0^2,$$
$$2ma_2s_2 = mv_2^2 - mv_1^2,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$2ma_ns_n = mv_n^2 - mv^2_{n-1},$$

and adding, writing F_1 for m_1a_1, etc., we have

$$\Sigma Fs = \tfrac{1}{2}mv_n^2 - \tfrac{1}{2}mv_0^2. \tag{31}$$

But ΣFs is the total work done by the acting forces, and $\frac{1}{2}mv_n^2 - \frac{1}{2}mv_0^2$ is the gain in kinetic energy. Hence more generally, the work done on a free body by the forces which act on it in any path is equal to the kinetic energy gained by the body.

If the work done is positive, the kinetic energy increases and the work done is equal to the gain in kinetic energy. If the work is negative, the kinetic energy diminishes and the work done is equal to the loss of kinetic energy. The relation expressed in the above formula was first given by Huygens.

83. Quantity of Motion. — The important place which kinetic energy holds among the fundamental quantities of mechanics was recognized by Leibnitz. He gave to the function mv^2 the name *vis viva*, by which name it was for a long time known. He was so impressed with its important relations that he considered it as measuring the quantity of motion of the body, in distinction from the view held by Descartes and Newton, who defined momentum as the measure of quantity of motion. Quantity of motion is so unintelligible a term that the controversy which arose about these

definitions could never be brought to a satisfactory issue, but it at least directed attention to the resemblances and differences of the two quantities.

They are both alike in being functions of the mass and the velocity and in expressing the relations of the body to some force which may be conceived as having acted to produce the motion, or which, by acting in opposition to the motion, may bring the body to rest. They differ in this, that the momentum expresses the relations of the velocity to the force in terms of the time during which the body will move until it comes to rest; while the *vis viva* expresses this relation in terms of the distance through which the body will move before it comes to rest. If bodies having different velocities travel against equal constant forces until they come to rest, the times during which they will be moving will be proportional to their momenta, while the distances they will traverse will be proportional to their *vis vivas*.

84. Work Done by an Impulse. — When a body is acted on by an impulsive force (§ 77), the work done by the force, in the short distance through which the body moves while acted on by force, cannot be calculated directly. It will, however, be equal to the increase in the kinetic energy of the body, or to

$$\tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2.$$

This quantity may be written in the form

$$\tfrac{1}{2}m\,(v - v_0)\,(v + v_0),$$

and $m\,(v - v_0)$ equals J, the impulse, by Equation 29. Hence the work done by an impulse J is equal to

$$J\frac{v + v_0}{2}. \tag{32}$$

Sometimes the distance s is known, through which the body moves while the impulsive force is acting. In such a case we may find an average value of the force from the formula

$$F = \frac{\tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2}{s}.$$

85. Work Done by a System of Forces. — It has been shown in another connection (§ 45) that when several forces act on a body, the work that is done when the body moves through any path is equal to that which would be done by the resultant of the forces in the same motion. Therefore if several forces act on a body which is otherwise free so that the body moves in the direction of their resultant, the work which is done, and the kinetic energy

which is gained, will be determined by the magnitude of the resultant and the extent of the motion.

Even if the body is constrained by some condition so that it is not entirely free to move in the direction of the resultant of the applied forces, it is always possible to represent the constraint by the introduction of an additional force of the appropriate magnitude and direction. In actual experience these additional forces are generally the reactions of some surface over which the body is constrained to move, or of some connecting piece by which it is prevented from moving except in certain paths. For instance, if the body is a smooth ball, rolling about in a smooth bowl, the applied force is the weight of the ball. The constraint is the condition that the ball shall always touch the bowl, and the equivalent force is the reaction of the bowl against the ball, which is a force always normal to its surface. Similarly if the body is a pendulum bob, the applied force is its weight, the constraint is the condition that the cord shall always be stretched, and the equivalent force is the tension in the cord. In each of these cases the body moves under the action of the resultant of these forces as if it were a free body.

The result of work done on a free body by a combination of forces is a change in its kinetic energy. The magnitude of this change is equal to the work done by the resultant and is positive or negative according as the resultant does positive or negative work. If the body starts from rest under the action of the given forces, it will move in the direction of their resultant, and the kinetic energy of the body will increase.

We may state the same conclusion in another way, by remembering that forces do positive or negative work, according as their components along the line of motion are in the direction of motion or are in the opposite direction; and that the component of the resultant in the line of motion is equal to the algebraic sum of the positive and negative components of the several forces. We may therefore say that the work done by the forces which do positive work is equal to the work done against the forces which do negative work added to the kinetic energy gained by the body. If we use W_p and W_n to represent the quantities of positive and of negative work done, and use K_1 and K_0 to represent the final and initial values of the kinetic energy, we may express this relation in the formula

$$W_p - W_n = K_1 - K_0. \tag{33}$$

In the special case in which the forces are in equilibrium, so that there is no resultant, the kinetic energy of the body does not change. In this case there is no work done on the body by the system of forces, considered as a whole, and the work done by the forces which do positive work is numerically equal to the work done by those which do negative work.

86. Potential Energy. — There are very many cases in nature in which the forces which act on a body, or at least those forces which are considered in the study of the body's motion, depend only on the position of the body relative to other bodies which act on it. Such forces may, for the present, be called *positional forces.* In case the bodies upon which they depend remain fixed, the positional forces which act upon a moving body, as it traverses any path, will be the same in whichever direction it is moving, and will not depend upon any peculiarities of its motion. The work which is done on the body by those forces, when it traverses any path, is therefore numerically equal and of opposite sign to the work which is done upon it when it traverses the same path in the opposite direction.

When a body is displaced in such a way that negative work has been done upon it by positional forces, it is said to have acquired *potential energy.* The potential energy thus acquired is measured by the work done against these positional forces, or by the work which they will do if the body is abandoned to their action and retraces its path in the opposite direction.

If the forces against which the work represented by W_n in Equation 33 is done are positional, the body acquires thereby potential energy $P_1 - P_0 = W_n$, representing by P_1 and P_0 the final and initial values of the potential energy. We then have

$$W_p = P_1 - P_0 + K_1 - K_0. \tag{34}$$

Other forces often act upon a body which do not depend solely on its position. Indeed, in actual mechanical operations, such forces always arise, though they are often ignored in formal problems. When they exist, it is found that they always depend upon the motion of the body. They may therefore, for the present, be called *motional forces.* A peculiarity which all motional forces possess in common is that they are always directed oppositely to the direction of the body's motion, so that the work which they do is negative. Work done against such forces does not give to the body potential energy, for the body, if left to itself, will not

be set in motion or be made to retrace its path by these forces. If some of the work represented by W_n in Equation 33 is done against motional forces, this does not appear as potential energy. Representing it by W_m, and the work done against positional forces again by $P_1 - P_0$ we have

$$W_p = W_m + P_1 - P_0 + K_1 - K_0. \tag{35}$$

87. Conservation of Mechanical Energy. — These positional and motional forces are commonly called *conservative* and *non-conservative* forces respectively. The reason for these names will be seen if we consider these forces with respect to the way in which they affect the mechanical energy of the body upon which they act. If the only forces which are acting are positional, the work done by those forces which do positive work is equal to the sum of the potential energy and the kinetic energy gained by the body. In the special case in which no positive work is done, the sum of the kinetic energy and the potential energy of the body remains constant, or is conserved. In this case $W_m = 0$ in Equation 35 from the condition that the only forces acting are positional, and $W_p = 0$ because by hypothesis no positive work is done. The law of conservation is then expressed by the relation

$$P_1 + K_1 = P_0 + K_0.$$

If some of the forces which act on the body are motional, the work done against them has no equivalent in potential energy gained by the body. In this case, therefore, if there are no forces doing positive work, the sum of the potential energy and the kinetic energy of a body does not remain constant, or is not conserved. In all cases in which motional forces occur, the sum of the potential energy and kinetic energy of the body continually diminishes. In this case W_m in Equation 35 does not vanish, and when $W_p = 0$, we have

$$P_1 + K_1 = P_0 + K_0 - W_m;$$

the sum $P_1 + K_1$ is less as W_m is greater.

88. Energy Criterion of Equilibrium. — When a system of bodies, initially at rest, possesses potential energy and is not acted on by external forces, the forces to which the potential energy is due will set the bodies in motion in such directions as to diminish the potential energy. Motions in directions such that by reason of them the potential energy increases are not spontaneous or natural, and will not be set up by the forces of the system (§ 29).

A body so situated that any conceivable displacement of it, consistent with the circumstances of its position, will result in an increase in its potential energy, will return toward its original position if it is slightly displaced from it. Such a body is said to be in *stable equilibrium*. When the displacement in any direction results in no change in the potential energy, the body will neither return nor proceed further if it is slightly displaced in that direction. It is then in *neutral equilibrium* with respect to motion in that direction. A body which is in equilibrium, but so situated that a displacement in any direction will result in a decrease in its potential energy, is in *unstable equilibrium* with respect to motion in that direction.

These cases are illustrated severally by a cone standing on its base, lying on its side, and balanced on its apex, on a flat horizontal table.

EXAMPLES, VII

1. *Show that when a force acts on a free body, the force is equal to the rate at which the kinetic energy changes as the body traverses space; or to the space-rate of change of kinetic energy.*

From Equation 31 we have

$$\Sigma Fs = \tfrac{1}{2}mv_n^2 - \tfrac{1}{2}mv_0^2.$$

If the space traversed is so small that we may consider the force acting in it as constant, we shall have from this

$$F = \frac{\tfrac{1}{2}mv_n^2 - \tfrac{1}{2}mv_0^2}{s},$$

and this is manifestly the rate of change of kinetic energy with respect to space. It may otherwise be stated to be the change in kinetic energy per unit of length.

2. *What constant force applied to a body of mass m to lift it from rest, will raise it s feet in t seconds?*

The applied force F acts against the weight of the body, mg, so that the acceleration a is given by $ma = F - mg$. But $2s = at^2$, so that $F = mg + \frac{2ms}{t^2}$.

3. *If a body is ascending, and passes a certain point with the velocity v, what constant force must be applied to it, as it moves through the distance s, to give it the velocity v'?*

To change the kinetic energy from $\frac{1}{2}mv^2$ to $\frac{1}{2}mv'^2$, work equal to $\frac{1}{2}mv'^2 - \frac{1}{2}mv^2$ must be done. The force doing the work is the difference between the applied force F and the weight mg. Setting the work done equal to the kinetic energy gained we have

$$F = \frac{\tfrac{1}{2}mv'^2 - \tfrac{1}{2}mv^2}{s} + mg.$$

When the force applied is such that the velocity does not change, it equals the weight.

4. **The Ballistic Pendulum.** — *A heavy block of wood is suspended so as to swing freely around a horizontal axis. A bullet is shot into it. What observations must be made to determine the velocity of the bullet?*

The bullet gives up its momentum to the block, so that if we know the masses m and M of the bullet and of the block, and can determine the velocity V of the block, we can get the velocity v of the bullet from the formula $mv = MV$, in which we neglect the small change in the mass of the block which the mass of the bullet will make. We may get V by observing the height h to which the block rises at the end of its swing. Equating kinetic and potential energy, we have $\frac{1}{2}MV^2 = Mgh$.

It is better to allow the block to swing as a pendulum, and to determine its period T. Then since its motion is approximately simple harmonic, the velocity at the mid-point of its swing is given by $V = \frac{2\pi a}{T}$, in which a is the amplitude, or extent of swing from the mid-point. This can be measured for each experiment, and the velocity thus determined.

The kinetic energy of the bullet is much greater than that of the block. Most of the bullet's energy is spent in doing work against non-conservative forces as it penetrates the block. The energy of the bullet is $\frac{1}{2}mv^2$; substituting for v from the equation of momentum, it is given by $\frac{1}{2}\frac{M^2V^2}{m} = \frac{M}{m} \cdot \frac{1}{2}MV^2$, so that the kinetic energy of the bullet is greater than that of the block in the ratio of the masses of the block and of the bullet. This relation ceases to hold when the bullet's mass is an appreciable fraction of the mass of the block.

DYNAMICS OF EXTENDED BODIES

Center of Mass

89. Body and Particle. — In the original statement of Newton's laws of motion, and in the use made of them, the motions are described as those of a body. Ordinarily the word *body* means a mass of matter occupying a finite volume; and if this meaning is accepted, there is an element of uncertainty in the application of the laws of motion, owing to the fact that when a body moves, its parts are not generally moving with equal velocities and accelerations. Those bodies to which the laws of motion apply unambiguously are bodies so small that the motions of their different parts are not perceptibly different. Such bodies are called *particles*. In certain circumstances extended bodies move in such a way that the motions of all their parts are alike or nearly alike, and can be represented either exactly, as in the case of a falling body, or approximately, as in the case of a swinging pendulum bob, by the motion of a single point. In all such cases the body behaves, with reference to the laws of motion, like a particle.

Any set of particles whose motion is under consideration is called a *system* of particles. A body may be considered as a system of particles rigidly connected together.

It is convenient to be able to reduce the description of the motion of a body or of a system of particles, as far as possible, to that of the motion of a single particle. This can be done by the help of the conception of the center of mass.

90. Center of Mass. — We may define the *center of mass* of two particles, as the point on the line which joins them (Fig. 69), which divides that line into segments inversely proportional to the masses at their extremities. If m_1 and m_2 represent the masses, d_1 and d_2 the segments of the line joining them, this definition gives

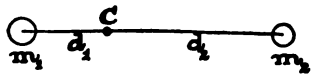


Fig. 69.

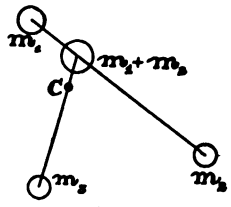


Fig. 70.

$$m_1 d_1 = m_2 d_2. \tag{36}$$

The center of mass of three particles (Fig. 70) is obtained by first locating the center of mass of two of them, supposing a mass equal to the sum of the two masses placed at that point, and then determining the center of mass between this imaginary mass and the third particle. The center of mass of a larger number of particles is obtained by an extension of the process just described.

We may define the center of mass in another way, which is peculiarly convenient when we come to examine its properties. To do this we suppose the position of each particle given by means of its coördinates x, y, z, referred to rectangular axes. Designating those giving the positions of different particles by subscripts, and using ξ, η, ζ to express the coördinates of the center of mass, we define them by the formulas

$$\left.\begin{aligned} \xi &= \frac{m_1 x_1 + m_2 x_2 + \cdots m_n x_n}{m_1 + m_2 + \cdots m_n}, \\ \eta &= \frac{m_1 y_1 + m_2 y_2 + \cdots m_n y_n}{m_1 + m_2 + \cdots m_n}, \\ \zeta &= \frac{m_1 z_1 + m_2 z_2 + \cdots m_n z_n}{m_1 + m_2 + \cdots m_n}. \end{aligned}\right\} \tag{37}$$

If we use $M = m_1 + m_2 + \ldots m_n$ to express the mass of the system, these equations may be written

$$M\xi = \Sigma mx, \quad M\eta = \Sigma my, \quad M\zeta = \Sigma mz. \tag{38}$$

To show that this definition of the center of mass is equivalent to the first one given, it will be sufficient to consider the simple case in which the particles lie in a plane, say the xy-plane.

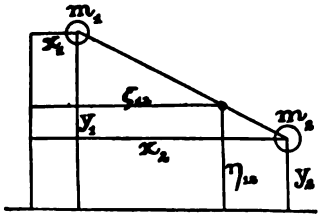


Fig. 71.

The center of mass of the two particles m_1, m_2 (Fig. 71) will be given by

$$(m_1 + m_2)\,\xi = m_1x_1 + m_2x_2,$$
$$(m_1 + m_2)\,\eta = m_1y_1 + m_2y_2.$$

From these equations we get

$$m_1(\xi - x_1) = m_2(x_2 - \xi),$$
$$m_1(\eta - y_1) = m_2(y_2 - \eta),$$

and these equations are also obtained by projecting on the axes the segments of the line joining the particles made by the center of mass, according to the first definition. The two definitions therefore designate the same point in this case, and the proof may readily be extended to take in three or any number of particles.

It might be thought that the choice of a different set of axes of reference might lead to the designation of a different point for the center of mass; but more extended analysis shows that this is not the case. The designation of the center of mass depends only upon the masses and distribution of the particles and is independent of the position of the axes.

If the origin of coördinates is taken at the center of mass, ξ, η, ζ vanish and we have $\Sigma mx = 0$, $\Sigma my = 0$, $\Sigma mz = 0$.

91. Velocity of the Center of Mass.— The assignment of the name center of mass to the point defined by the equations of the last section is justified by the fact that the motions of that point depend in a simple and systematic way upon the motions of the individual particles.

To find the velocity of the center of mass we suppose that the particles of the system move during a short time t, in which we can consider their velocities constant, from positions indicated by unprimed letters to positions indicated by primed letters. The x-coördinates of the center of mass at the beginning and end of the time t will be given by

$$M\xi = m_1x_1 + m_2x_2 + \cdots m_nx_n,$$
$$M\xi' = m_1x_1' + m_2x_2' + \cdots m_nx_n',$$

and the y- and z-coördinates will be given by similar equations. Subtracting the first equation of each set from the second, and dividing by t we obtain

$$M\frac{\xi' - \xi}{t} = m_1\frac{x_1' - x_1}{t} + m_2\frac{x_2' - x_2}{t} + \cdots m_n\frac{x_n' - x_n}{t},$$
$$M\frac{\eta' - \eta}{t} = m_1\frac{y_1' - y_1}{t} + m_2\frac{y_2' - y_2}{t} + \cdots m_n\frac{y_n' - y_n}{t},$$
$$M\frac{\zeta' - \zeta}{t} = m_1\frac{z_1' - z_1}{t} + m_2\frac{z_2' - z_2}{t} + \cdots m_n\frac{z_n' - z_n}{t}.$$

The factors $\frac{\xi' - \xi}{t}, \frac{\eta' - \eta}{t}, \frac{\zeta' - \zeta}{t}$ are the components of the velocity of the center of mass. The terms in any one column on the right are the components of the momentum of one of the particles, and the sums on the right are the sums of those components. We therefore obtain the components of the velocity of the center of mass by adding the components of the momenta of the several particles and dividing by the sum of the masses.

The velocity of the center of mass, obtained from the three component velocities, is equal to the square root of the sum of their squares. The square root of the sum of the squares of the three component momenta is the resultant of the momenta of the particles in the system. We therefore obtain the velocity of the center of mass by the following rule:

The velocity of the center of mass is obtained by dividing the resultant momentum of the system by the sum of the masses.

Using U, V, W, u_1, v_1, w_1, etc., for the component velocities of the center of mass and of the individual particles, the velocity of the center of mass is given by the equations

$$MU = \Sigma mu, \quad MV = \Sigma mv, \quad MW = \Sigma mw. \tag{39}$$

If the origin of coördinates is at the center of mass and moves with it, U, V, W vanish and we have $\Sigma mu = 0$, $\Sigma mv = 0$, $\Sigma mw = 0$.

92. Acceleration of the Center of Mass. — If any of the particles of a system are acted on by forces, so that the motion of the particles is accelerated, the motion of the center of mass will in general be accelerated also. To find the acceleration we suppose that the velocities of the particles change during a time t, so short that the accelerations during it may be considered constant, from values indicated by unprimed letters to values indicated by primed letters. Writing out the equations for the component velocities along the x-axis at the beginning and end of the time t we have

$$\begin{aligned} MU &= m_1u_1 + m_2u_2 + \ldots m_nu_n, \\ MU' &= m_1u_1' + m_2u_2' + \ldots m_nu_n', \end{aligned}$$

and similar equations hold for the component velocities along the other axes.

Subtracting the first equation of each set from the second and dividing by t we have

$$M\frac{U'-U}{t} = m_1\frac{u_1'-u_1}{t} + m_2\frac{u_2'-u_2}{t} + \ldots m_n\frac{u_n'-u_n}{t},$$

$$M\frac{V'-V}{t} = m_1\frac{v_1'-v_1}{t} + m_2\frac{v_2'-v_2}{t} + \ldots m_n\frac{v_n'-v_n}{t},$$

$$M\frac{W'-W}{t} = m_1\frac{w_1'-w_1}{t} + m_2\frac{w_2'-w_2}{t} + \ldots m_n\frac{w_n'-w_n}{t}.$$

The factors $\frac{U'-U}{t}$, $\frac{V'-V}{t}$, $\frac{W'-W}{t}$ are the component accelerations of the center of mass. The terms in any column on the right are the products of the mass of a particle by its component accelerations, or are equal to the components along the axes of the force acting on the particle.

Using A, B, C, a_1, b_1, c_1, etc., to represent the component accelerations of the center of mass and of the individual particles, the acceleration of the center of mass is given by the equations

$$MA = \Sigma ma,\ MB = \Sigma mb,\ MC = \Sigma mc. \qquad (40)$$

We obtain from these equations the following rule:

The acceleration of the center of mass is obtained by dividing the resultant of the forces acting on the particles of the system by the sum of the masses.

Since the sum of the masses multiplied by the acceleration of the center of mass is equal to the resultant of all the forces acting on the system, this rule may be otherwise stated by saying that the center of mass moves as if all the masses of the system were concentrated in it, and were there acted on by a force equal to the resultant of all the forces.

This law has been tacitly assumed in all the experiments in which we have used moving weights to illustrate the simple mechanical laws. We have treated the weight of the body in each case as if it were a single force, and have treated the body as a mass situated at a point. The real forces were of course the weights of the different parts of the body. Their resultant was the force which we used as the weight of the body, and its point of application, whose motion was studied, was the center of mass.

The forces which act upon a system of bodies arise either from the action of bodies outside the system, or from the interaction of bodies in the system. Such forces are called *external* and *internal* forces respectively. It is of considerable importance to notice that the motion of the center of mass is not affected by the action of internal forces. For, by the third law of motion, these internal forces always occur as pairs of equal and opposite forces. The resultant of each such pair of forces is zero, and they therefore contribute nothing to the resultant force from which the motion of the center of mass is determined. If the center of mass is at any time at rest, it cannot be set in motion by the internal forces of the system, and will therefore remain at rest unless external forces act.

93. The Kinetic Energy of a System of Particles. — The kinetic energy of a system of moving particles may be expressed in such a way as to include in its expression the velocity of the center of mass. To form this expression we construct a diagram by drawing vectors from a point O to represent the velocities $v_1, v_2, \ldots v_n, V$ for the several particles and for the center of mass (Fig. 72), and from the terminal point of the vector V, representing the velocity of the center of mass, we draw vectors $u_1, u_2, \ldots u_n$ to the terminal points of the vectors $v_1, v_2, \ldots v_n$. These vectors, being the vector differences between the several velocities and the velocity of the center of mass, are the velocities of the several particles relative to the center of mass taken as origin. We shall designate the projections of $u_1, u_2 \ldots u_n$ on the line of V by $b_1, b_2 \ldots b_n$. Then from the triangle whose sides are v_1, u_1, and V we have

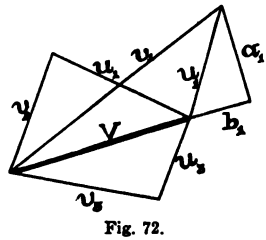


Fig. 72.

$$v_1^2 = V^2 + u_1^2 + 2\,b_1 V.$$

Multiplying this equation by $\frac{1}{2}m_1$, we have

$$\tfrac{1}{2}m_1 v_1^2 = \tfrac{1}{2}m_1 V^2 + \tfrac{1}{2}m_1 u_1^2 + m_1 b_1 V,$$

and similar equations hold for the other masses. Adding them we obtain on the left the sum of the kinetic energies of the several particles, or the kinetic energy K of the system. The first term on the right is $\frac{1}{2}MV^2$. The sum of the second terms is the sum of the kinetic energies of the several particles when moving with the velocities which they have relative to the center of mass. The sum of the last terms is

$$(m_1 b_1 + m_2 b_2 + \ldots m_n b_n)V$$

and equals zero, since the sum in parenthesis represents the resultant momentum of the system in the direction of V referred to the center of mass as origin (§ 91). We therefore obtain for the kinetic energy the equation

$$K = \tfrac{1}{2}MV^2 + \Sigma\,\tfrac{1}{2}mu^2; \qquad (41)$$

and can give for it the following rule:

The kinetic energy of a system of particles is the sum of the kinetic energies of a mass equal to the sum of the masses of the particles and moving with the velocity of the center of mass, and of the several particles moving with the velocities which they have relative to the center of mass.

If the center of mass is an origin from which the velocities of the parts of the system may be determined, the kinetic energy of the system will be the sum of the kinetic energies of the parts determined from these velocities. If, besides the motions of the parts relative to the center of mass, the system as a whole has a motion relative to some external point, the kinetic energy of the system relative to that point is obtained by adding to the kinetic energy already obtained an additional amount equal to the kinetic energy obtained by supposing all the system concentrated at the center of mass, and moving with the velocity of the center of mass relative to the external point.

It follows from this statement that the kinetic energy of a system will never be zero unless each part of the system is at rest. The system as a whole may have no momentum even though its parts are moving, provided its center of mass is at rest, but it will always have kinetic energy unless each of its parts separately is at rest. The reason for this difference lies in this: that momentum is a vector and that consequently two momenta in opposite directions may be added so as to be equal to no momentum; while kinetic energy is not a vector, is always a positive quantity, and so when two kinetic energies are added together the result is always positive.

EXAMPLES, VIII

1. **Collision of Bodies.** — When the system consists of two bodies, and these bodies interact by collision, the forces called out by the collision are internal forces, and the velocity of the center of mass remains unchanged. In the simplest case, in which the bodies are spheres, whose centers are moving along the same line, the collision gives rise to forces which are in the same line. Calling the masses of the spheres m_1 and m_2, their velocities before collision u_1 and u_2, their velocities after collision v_1 and v_2, and the velocity of the center of mass v, we must have

$$(m_1 + m_2)\, v = m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2. \tag{42}$$

This equation is not sufficient to enable us to find the velocities of the spheres after collision. Another equation is needed, which is supplied by the use of an experimental law discovered by Newton, that when two bodies collide in the way we have assumed, their relative velocity changes in a constant ratio. Stated in symbols this law gives

$$e\,(u_1 - u_2) = v_2 - v_1, \tag{43}$$

in which e is experimentally a constant number. It is called the *coefficient of restitution*. Its value may be as great as 1 in the ideal case of perfectly elastic bodies, but in real bodies it is always less than 1, and in many cases of soft, adherent bodies, it is not perceptibly different from 0.

From these equations we get

$$v_1 = v - \frac{m_2}{m_1 + m_2}\, e\,(u_1 - u_2), \qquad v_2 = v + \frac{m_1}{m_1 + m_2}\, e\,(u_1 - u_2).$$

If $e = 0$, the spheres acquire the same velocity, $v_1 = v_2 = v$, the velocity of the center of mass. In the ideal case in which $e = 1$, we have

$$v_1 = \frac{(m_1 - m_2)\, u_1 + 2\, m_2u_2}{m_1 + m_2}, \qquad v_2 = \frac{2\, m_1u_1 - (m_1 - m_2)\, u_2}{m_1 + m_2}. \tag{44}$$

In experimenting on collisions, the velocity u_2 is often 0; in this case we have

$$v_1 = \frac{(m_1 - m_2)\, u_1}{m_1 + m_2}, \qquad v_2 = \frac{2m_1 u_1}{m_1 + m_2}.$$

If in this case we further take the two spheres equal, we have

$$v_1 = 0, \quad v_2 = u_1,$$

so that the spheres exchange velocities.

If $u_2 = 0$, and m_2 is very great, so that it can be taken as infinitely great, we have from the general equation

$$v_1 = -eu_1, \quad v_2 = 0.$$

This result suggests a way to determine e. A sphere may be dropped from a known height h on a heavy slab of marble or steel, and the height k to which it rebounds may be measured. The value of u_1 is given by $u_1 = \sqrt{2gh}$, and of v_1 by $v_1 = \sqrt{2gk}$, so that $e = \sqrt{\frac{k}{h}}$.

The kinetic energy of the spheres before collision is given in terms of the velocity of the center of mass by

$$\tfrac{1}{2}(m_1 + m_2)\, v^2 + \tfrac{1}{2}m_1 (u_1 - v)^2 + \tfrac{1}{2}m_2 (v - u_2)^2.$$

After collision the kinetic energy is

$$\tfrac{1}{2}(m_1 + m_2)\, v^2 + \tfrac{1}{2}m_1 (v - v_1)^2 + \tfrac{1}{2}m_2 (v_2 - v)^2.$$

From the fundamental equations it is easy to show that $v - v_1 = e\,(u_1 - v)$; $v_2 - v = e\,(v - u_2)$, and making these substitutions in the second equation and subtracting it from the first, we get for the kinetic energy lost by collision

$$\frac{1 - e^2}{2}\,(m_1 (u_1 - v)^2 + m_2 (v - u_2)^2).$$

By substituting the value of v in terms of u_1 and u_2 this is easily reduced to

$$\frac{m_1 m_2}{m_1 + m_2} \cdot \frac{1 - e^2}{2}\,(u_1 - u_2)^2.$$

In the ideal case in which $e = 1$ there is no kinetic energy lost by collision. In all other cases kinetic energy is lost.

Uniplanar Motion of Rigid Bodies

94. Motion of Extended Bodies. — In discussing the motion of bodies which are too large to be treated as particles, it is convenient to limit ourselves to motion in a plane. Motion in three dimensions is so complicated that even its fundamental characteristics cannot be demonstrated by elementary methods. In uniplanar motion, the parts of the body move in the same plane or in parallel planes. The formal object of our study will therefore be a plane figure moving in its own plane, and carrying masses which are placed at some or all of the points of the figure and rigidly attached to each other.

When a plane figure moves so that all its points move in similar and equal paths, its motion is called a *translation*. A translation

is effected when each point of the body undergoes the same displacement. In Fig. 73 is represented the translation of the figure carrying the line AB to a new position marked by $A'B'$.

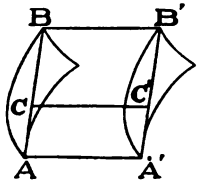


Fig. 73.

When a plane figure moves so that each point of it describes an arc of a circle around a single point of the plane, which is the common center of all the circular paths described, its motion is called a *rotation*. In Fig. 74 is represented the rotation of the figure carrying the line AB about a point C to a new position marked by $A'B'$.

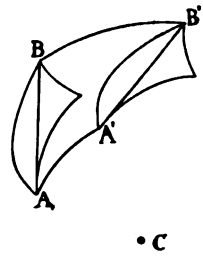


Fig. 74.

Any displacement of a plane figure in a plane may be obtained by the combination of a translation and a rotation. In Fig. 75, in which the operation by which this is done is described, the figure is represented by a single line in it, AB, as may obviously be done in all cases. We select a point C either in the figure or in an extension of the figure, supposed to move with it, as the center about which rotation is to be effected. This point is transferred, by a translation of the figure, from its original position to that which it occupies in the position given the figure by the displacement. The figure is then rotated around this point C' as center into the position which is given it by the displacement. Since a translation in any direction transfers any straight line AB drawn in the figure to a new position $A''B''$, in which that line is parallel with its original position, it is plain that the rotation involved in this operation will be the same, that is, will turn the straight line through the same angle, whatever point in the figure is taken as the center of rotation. The magnitude and the direction of the translation will in general depend upon the point chosen as the center of rotation.

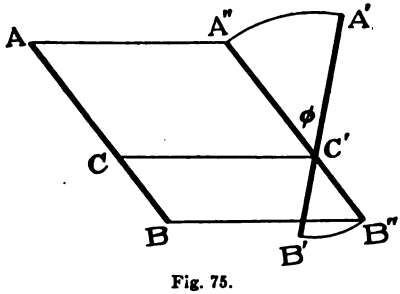


Fig. 75.

In particular, a point may always be chosen as the center of rotation such that no translation is needed to effect its displacement, which can be effected by a simple rotation around that point as center. Let the line AB (Fig. 76) represent the initial, and the line $A'B'$ the final position of the line representing the figure.

Draw the lines AA', BB', bisect them and erect perpendiculars at the points of bisection. These will meet at the point C. This point is then such that a rotation of AB around it as center will bring it to the position $A'B'$. For by construction we have $CA = CA'$, $CB = CB'$; and therefore the triangle CAB equal to the triangle $CA'B'$, and the angle ACB equal to the angle $A'CB'$. If we add to each of these the angle BCA' we have the angle ACA' equal to the angle BCB'. Therefore the points A' and B' are at the same distances from C as A and B respectively and have been reached by the same angular rotation, so that pure rotation around C has transferred AB to $A'B'$.

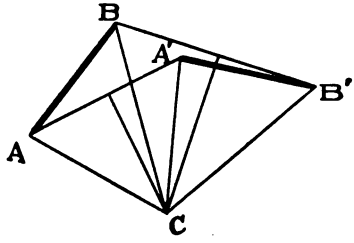


Fig. 76.

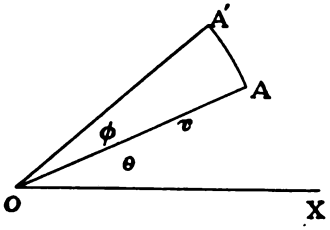


Fig. 77.

95. Rotation. — In order to describe the rotation of a body and its dependence upon the forces which act on the body, it is convenient at this point to introduce a method of describing the motion of the body in terms of angular magnitudes.

Consider a plane figure (Fig. 77) free to rotate about a point O. Draw a line OA from this point to any definite point of the figure. This line merely marks a row of points of the figure. From the same point or center of rotation, draw another line OX of indefinite length, which is fixed in the plane in space and serves as an axis of reference for the measurement of angles. The position of the figure in the plane is then determined, at any instant, by the angle θ between this axis and the line first drawn in the figure. When rotation takes place, the angle between these lines changes, or the figure, as determined by the line drawn in it, rotates through a certain angle ϕ. This angle is called the *angular displacement.*

Each of the points of the figure moves through an arc of a circle whose radius r is its distance from the center of rotation. If the angle in which the angular displacement is given is measured in radians, the distance s traversed by each point of the figure, in the circular arc which it describes, is equal to its distance from the center multiplied by the angular displacement, or $s = \phi r$.

96. Angular Velocity. — If the figure is rotating uniformly, so that its angular displacements in equal times are equal, it has a constant *angular velocity*, measured by the angular displacement

which occurs in a given time, divided by that time. Representing angular velocity by ω, the formula defining it is

$$\omega = \frac{\phi}{t} . \tag{45}$$

If the angular displacement is not the same for any equal intervals of time, the angular velocity of the figure is variable. Its value at any instant is the limit of the ratio of the angular displacement which occurs to the time in which it occurs, as the time and so also the angular displacement approach zero. The unit of angular velocity is the angular velocity of a uniformly rotating line which sweeps out one radian in one second.

From the relation already described between the distance actually traversed by a point in the figure and the angular displacement, it is plain that the velocity v of any point is equal to the angular velocity of the figure multiplied by the distance of that point from the center of rotation, or $v = \omega r$.

97. Angular Acceleration. — When the angular velocity of the figure changes at a uniform rate, so that equal changes of angular velocity occur in equal times, the figure is said to have a constant *angular acceleration*. This constant angular acceleration is the ratio of the change in angular velocity to the time in which that change occurs. If we represent angular acceleration by α, and the initial and final values of the angular velocity by ω_0 and ω, the formula defining angular acceleration is

$$\alpha = \frac{\omega - \omega_0}{t} . \tag{46}$$

The unit angular acceleration is the angular acceleration of a rotating line whose angular velocity increases uniformly by one unit of angular velocity in one second.

If we consider the motion of any point of the body, it is plain that the rate at which the numerical value of its velocity changes is equal to the angular acceleration multiplied by the distance of that point from the center of rotation, so that $\frac{v - v_0}{t} = \alpha r$.

98. Constant Angular Acceleration. — If a body is in rotation with a constant angular acceleration α for a time t, it will in that time rotate through some angle. If the initial angular velocity is ω_0 and the final angular velocity ω, and if the angular velocity increases uniformly, the average angular velocity of the body during the time t will be $\frac{\omega + \omega_0}{2}$, and the angle ϕ described in that time will be given by

$$\phi = \frac{\omega + \omega_0}{2} t. \tag{47a}$$

Using the equation $\alpha = \dfrac{\omega - \omega_0}{t}$ and eliminating ω, we have

$$\phi = \omega_0 t + \tfrac{1}{2}\alpha t^2. \tag{47b}$$

By eliminating t from these equations we have also

$$\omega^2 - \omega_0^2 = 2\alpha\phi. \tag{47c}$$

If the rotation begins at the beginning of the time t, we have $\omega_0 = 0$ and the equations become

$$\left.\begin{aligned} \omega &= \alpha t, \\ \phi &= \frac{\omega t}{2} = \frac{1}{2}\alpha t^2. \\ \omega^2 &= 2\alpha\phi. \end{aligned}\right\} \tag{48}$$

These equations connecting the angular displacement, velocity, and acceleration with one another and with the elapsed time are analogous to those obtained for the motion of a particle moving in a straight line with a constant acceleration.

99. Kinetic Energy of a Rotating Body. — A plane figure rotating around a point, and carrying masses at some or all of its points, is similar to a body rotating around an axis passing through the center of rotation and perpendicular to the plane of the figure. If such a body rotates with an angular velocity ω, it possesses kinetic energy. The magnitude of its kinetic energy is the sum of the kinetic energies of the masses which constitute it. The velocity of one of these masses, whose distance from the center of rotation is r, is $r\omega$, and its kinetic energy is $\frac{1}{2}mr^2\omega^2$. By adding together the expressions such as this for the kinetic energies of the different masses which make up the body, we obtain for the kinetic energy of the whole body $\dfrac{\omega^2}{2}\Sigma mr^2$. The factor Σmr^2, which manifestly depends only on the masses which make up the body and on their respective distances from the axis of rotation, is called the *moment of inertia* of the body. With respect to angular motions of the body, the moment of inertia plays the part of the mass of the body with respect to linear motions.

100. Effect of Force on a Rotating Body. — When a body free to rotate is acted on by a force, it will in general have an angular acceleration. The relation between the force and the angular acceleration which it imparts to the body may be found from the general relation between the work of a force and the kinetic energy produced by it. To do this most simply we may consider a heavy wheel

(Fig. 78), whose moment of inertia is represented by I, mounted on a cylindrical axle whose radius is p, and free to rotate about the line which is the axis of this cylinder. If one end of a flexible cord is attached to the axle, and if the cord is then wound around the axle, a force applied to the free end of the cord will set the wheel in rotation. Furthermore, as the cord unwinds, this force will always be similarly applied to the axle. Let us compare the work done with the kinetic energy gained during a short interval of time in which a very small length s of the cord is drawn off the axle. The work done is $Fs = Fp\phi$, if ϕ is the angular displacement of the wheel, so that $s = p\phi$. The kinetic energy gained (§ 99) is $\frac{1}{2}I(\omega^2 - \omega_0^2)$, if ω and ω_0 are the final and initial values of the angular velocity. During the short time of the motion, the angular acceleration α may be considered constant, and hence (Equation 47c) $\omega^2 - \omega_0^2 = 2\alpha\phi$.

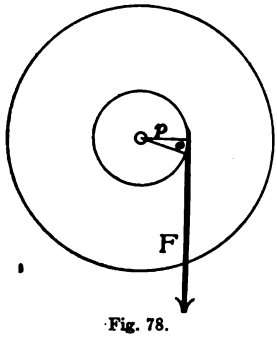


Fig. 78.

Using this in the equation for the kinetic energy and setting the work done equal to the kinetic energy gained, we have

$$Fp = I\alpha. \tag{49}$$

The product Fp is the moment of the force F about the axis of rotation (§ 18). When used in this way, as the measure of the tendency of the force to turn a body about an axis, it is often called the *torque*. The relation expressed by this equation may be stated by saying, that the moment of force which acts on a rotating body is equal to the moment of inertia of the body multiplied by the angular acceleration imparted by the force. This relation is analogous to the fundamental relation connecting force, mass, and acceleration in linear motion.

If several forces act on a rotating body at once, each of them will impart an angular acceleration proportional to its moment. These angular accelerations may be such as either to increase or to decrease the angular velocity of the body. Those which increase the angular velocity of the body may be considered positive, the others negative. The algebraic sum of all of them will be the angular acceleration of the body caused by all the forces. If we consider those moments of force positive which tend to turn the body in one sense, and those negative which tend to turn it in the opposite sense, we

may add the moments of force and their equivalents algebraically and obtain

$$F_1p_1 + F_2p_2 + \cdots F_np_n = I(\alpha_1 + \alpha_2 + \cdots \alpha_n)$$

or

$$\Sigma Fp = I\alpha. \tag{50}$$

The algebraic sum ΣFp of the moments of force is equivalent to a single moment of force.

If the angular acceleration of the body is zero, so that the body is in equilibrium, we obtain $\Sigma Fp = 0$ as the condition of equilibrium. This condition is the same as that found by experiment in § 19.

101. Effect of an Impulse on a Rotating Body. — If an impulsive force is applied to a body, which is free to turn about an axis, it will give the body an angular velocity about the axis. We can find a relation between the impulse and the angular velocity imparted to the body by the use of the relation discussed in the last section.

We suppose that the short time during which the impulsive force is applied is divided into many small parts $\tau_1, \tau_2, \ldots \tau_n$, each of them so short that the force which acts during it may be considered constant. We designate the angular velocity at the beginning of the first time interval by ω_0, and the angular velocities at the ends of the successive time intervals by $\omega_1, \omega_2, \ldots \omega_n$; the forces in the successive time intervals by $F_1, F_2, \ldots F_n$, the corresponding angular accelerations by $\alpha_1, \alpha_2, \ldots \alpha_n$. The direction of the forces will not change during the impulse, and p will be constant. We then have for each of the time intervals, from Equation 49,

$$F_1p = I\alpha_1 = I\frac{\omega_1 - \omega_0}{\tau_1},$$

$$F_2p = I\alpha_2 = I\frac{\omega_2 - \omega_1}{\tau_2},$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$F_np = I\alpha_n = I\frac{\omega_n - \omega_{n-1}}{\tau_n};$$

or by addition

$$p\Sigma F\tau = I(\omega_n - \omega_0). \tag{51}$$

The sum $\Sigma F\tau$ is the impulse, and the product $p\Sigma F\tau$ is the moment of the impulse. We name the product $I\omega$, the *angular momentum*. The formula states that the moment of impulse equals the increase in the angular momentum.

102. Effect of a Couple on a Free Body. — A couple applied to a body free to move in a plane will produce rotation around the center of mass. For, from the proposition stated in § 92, the acceleration of the center of mass depends upon the resultant of the forces which act on the body. Since the couple has no resultant, the center of mass will have no acceleration, and the motion of the body will therefore be a rotation around the center of mass. The moment of couple will be equal to the moment of inertia of the body about an axis through the center of mass multiplied by the angular acceleration.

If two equal impulses act on a free body in opposite directions along parallel lines, they form an *impulsive couple*. An impulsive couple gives the body angular momentum about an axis passing through its center of mass. The gain in angular momentum is equal to the impulsive couple.

103. Effect of a Force on a Free Body. — When a force is applied to a free body, its effect will be, in general, to impart acceleration to the center of mass and also to cause rotation around that center. This may be shown for the plane figure as follows (Fig. 79). Apply to the center of mass two forces each numerically equal to the given force, one of them parallel to the given force and in the same direction, the other opposite to it. These two forces being equal and opposite and applied at the same point, will have no effect on the motion of the body. They constitute, along with the given force, a system of three forces. According to the proposition of § 92 the application of the given force to the body will impart an acceleration to the center of mass. This acceleration may be considered as arising from the action of that one of the two forces which is applied to the center of mass and is parallel to and in the same direction as the given force. The other two forces may be considered as constituting a couple which will produce rotation around the center of mass.

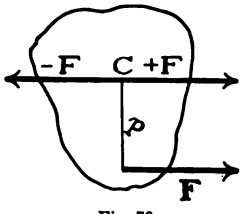


Fig. 79.

An impulse applied to a free body may be similarly analyzed into an impulse applied at the center of mass, and an impulsive couple. The impulse gives the body linear momentum; the impulsive couple gives it angular momentum about the center of mass.

Any number of forces applied in a plane containing the center of mass of the body, may similarly be analyzed into a set of forces

applied at the center of mass, and a set of couples. The forces combine to form a single resultant, and the couples are equivalent to a single resultant couple, and the motion of the body is the same as if it were acted on by this resultant force and resultant couple. The conditions of equilibrium of the body with respect to motion in this plane are that the resultant force and resultant couple should both vanish. That the forces applied to a body in general should be in equilibrium, the resultant forces and couples must vanish in each of three mutually perpendicular planes containing the center of mass.

104. Moments of Inertia. — To apply the formula for angular acceleration it is necessary to know the moment of inertia of the body which is set in rotation by the applied force. The moment of inertia is a function of the masses and their distances from the axis of rotation, and can be calculated in some cases approximately, by direct summation, if the masses and their distances from the axis are known. Thus an approximate value for the moment of inertia of a flywheel, in which most of the mass is collected in the rim, may be obtained by multiplying the mass of matter in the rim by the square of a mean radius of the rim. Such a method, however, imperfect as it is, can only be applied in some few cases.

The moments of inertia of bodies in which the density is uniform and which have a regular figure referred to some axis of symmetry in the figure, may be calculated by the methods of the integral calculus. We shall not attempt to obtain the formulas for any such cases, but shall content ourselves with the citation of some of those which are most generally useful.

Rectangular parallelepiped referred to a central axis perpendicular to one of its faces. Represent by a the length, by b the breadth, of the face perpendicular to the axis, and by M the mass of the body: then

$$I = M\frac{a^2 + b^2}{12}. \tag{52}$$

Cylinder referred to its axis of figure. Represent by b its diameter, and by M its mass: then

$$I = M\frac{b^2}{8}. \tag{53}$$

Cylinder referred to an axis through its center perpendicular to its axis of figure. Represent by a its length, by b its diameter, and by M its mass: then

$$I = M\left(\frac{a^2}{12} + \frac{b^2}{16}\right). \tag{54}$$

Sphere referred to a diameter. Represent by r its radius and by M its mass: then

$$I = \tfrac{2}{5}Mr^2. \tag{55}$$

Many combinations of these bodies may be built up to form the bodies on which experiments may be tried, and by the aid of an additional theorem, their moments of inertia may be calculated about the common axis of rotation of the composite body. This theorem is as follows: The moment of inertia of a body referred to any axis may be found by adding to the moment of inertia of the body referred to a parallel axis passing through its center of mass the product of its mass and the square of the distance between the axes. If we represent the moment of inertia referred to an axis passing through the center of mass by K, and the distance between that axis and another parallel to it by R, then the moment of inertia referred to the parallel axis is

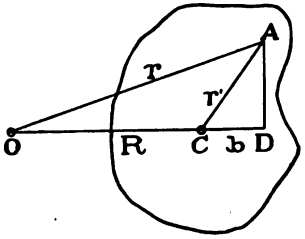


Fig. 80.

$$I = K + MR^2. \tag{56}$$

To prove this let us consider any body referred to the axes which stand perpendicularly to the paper (Fig. 80) at O and C. The axis at C passes through the center of mass. We suppose a particle of mass m to stand at the point A, of which the distance from the axis O is r, and from the axis C is r'. Then the moment of inertia of this particle referred to the axis O is mr^2, and that of the whole body is the sum of similar terms, or Σmr^2. But from the geometry of the triangle

$$r^2 = R^2 + r'^2 + 2bR,$$

if we represent by b the intercept on the side R produced between the point C and the perpendicular to that line from the point A. The moment of inertia I of the body about O is therefore

$$\Sigma mr^2 = R^2\Sigma m + \Sigma mr'^2 + 2R\Sigma mb.$$

The first term on the right is equal to MR^2; the second term is the moment of inertia K of the body about the axis passing through the center of mass; and the third term vanishes, because from the definition of the center of mass the sum $\Sigma mb = 0$. Hence

$$I = K + MR^2.$$

With this theorem and the obvious supplement to it that the moment of inertia of a composite body is the sum of the moments of inertia of its parts, we may obtain moments of inertia in most important cases.

If we divide the moment of inertia Σmr^2 of a body referred to any axis by its mass M we obtain a quantity which is evidently the square of a length. This length is called the *radius of gyration* of the body referred to the axis. Representing it by k we have

$$\Sigma mr^2 = Mk^2, \tag{57}$$

and are led to define the radius of gyration as the distance from the axis at which a mass equal to that of the body should be concentrated to have the same moment of inertia as that of the body.

105. The Physical Pendulum. — Any body which is free to rotate about a fixed horizontal axis and which swings back and forth in a vertical plane under the action of its own weight, is a

physical pendulum. As was explained in § 62, the time of oscillation of such a pendulum must be the same as that of a certain simple pendulum. The problem before us is to determine, from the characteristics of the physical pendulum, the length of the simple pendulum whose period of oscillation is the same as that of the physical pendulum. When the pendulum is not swinging, the line drawn (Fig. 81) from the axis of suspension to its center of gravity is vertical. When the pendulum swings, the extent of its swing at each instant is measured by the angle ϕ which this line drawn in the pendulum makes with the vertical. We designate the mass of the pendulum by M, and the distance of its center of gravity from the axis of suspension by R. The moment of force which acts upon the pendulum when its deviation is ϕ is its weight Mg applied at its center of gravity multiplied by the distance $R \sin \phi$ from the axis of suspension to the line of direction of the weight. This being so, the angular acceleration of the pendulum varies with ϕ and is given at the instant at which the deviation is ϕ by the formula $MgR \sin \phi = I\alpha$, or $\alpha = \frac{MgR}{I} \sin \phi$ (§ 100). The motion of the pendulum is therefore such that its angular acceleration is proportional to the sine of its deviation, and the factor $\frac{MgR}{I}$ is the factor of proportion.

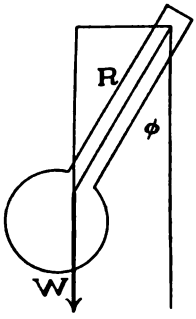


Fig. 81.

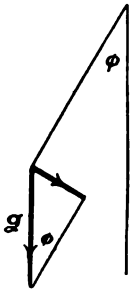


Fig. 82.

In the case of the simple pendulum (Fig. 82) the acceleration a of the bob, tangent to its circular path, is given for any deviation ϕ by $a = g \sin \phi$ (§ 63). Dividing by the length r of the pendulum and writing $\frac{a}{r} = \alpha$, the angular acceleration, we have $\alpha = \frac{g}{r} \sin \phi$. A comparison of this formula with the one obtained for the physical pendulum in the preceding paragraph shows that when the deviations of the two pendulums are the same, that is, when the values of ϕ are the same in both formulas, the accelerations will be the same and therefore the motions will be in every respect the same, if the factor of proportion $\frac{MgR}{I}$ is equal to the corresponding factor of proportion $\frac{g}{r}$. Setting these two quantities equal, we obtain

for the length of the simple pendulum which will swing in the same period as the physical pendulum the expression $r = \frac{I}{MR}$. The quantity I is the moment of inertia of the pendulum about its axis of suspension. The quantity MR is called the static moment.

The problem of finding the period of oscillation of the physical pendulum is thus reduced to the problem of finding the period of the equivalent simple pendulum. In § 61 we found that the period of the simple pendulum swinging in an infinitesimal arc is given by the formula $t = \pi\sqrt{\frac{r}{g}}$; and therefore the period of the equivalent physical pendulum swinging in an infinitesimal arc will be given by the formula

$$t = \pi\sqrt{\frac{I}{MRg}}. \tag{58}$$

We may therefore find the value of g by observing the period of a pendulum so constructed that its moment of inertia I and the position of its center of gravity can be calculated. Such a pendulum was first used by Huygens. From its use in an elaborate investigation by Borda it is often called Borda's pendulum.

106. The Reversible Pendulum. — In the previous section we have found that the physical pendulum will swing in the same period as a certain simple pendulum whose length is computed by the formula

$$r = \frac{I}{MR}. \tag{59}$$

This length of the equivalent simple pendulum is often called the *reduced length* of the physical pendulum. If a distance equal to it is measured off from the axis of suspension and perpendicular to it, through the center of gravity of the pendulum, the point at its further end is called the *center of oscillation* of the pendulum.

In the formula for r we may substitute for I (§ 104) its value in terms of the moment of inertia K of the pendulum about a parallel axis passing through its center of mass, and then eliminate M by introducing the radius of gyration k about the axis passing through the center of mass (§ 104). We thus obtain

$$r = \frac{K + MR^2}{MR} = \frac{k^2 + R^2}{R}. \tag{60}$$

The only quantity in this expression which can vary is R, and if it is constant, r is constant; so that the pendulum will have the

same period if swung from any point of a circle of radius R drawn in the plane of oscillation with the center of mass as center (Fig. 83).

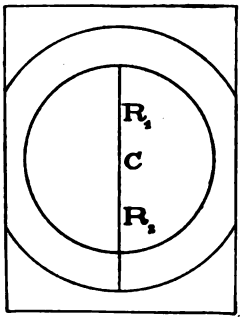


Fig. 83.

Furthermore, the equation in R is a quadratic and, for a given value of r and therefore of the period, R has two values, given by the equation

$$R = \frac{r}{2} \pm \sqrt{\frac{r^2 - 4k^2}{4}}.$$

Calling these values R_1 and R_2 and adding them we get

$$r = R_1 + R_2.$$

Two circles may therefore be drawn about the center of mass, of radii R_1 and R_2, and the pendulum will swing about any point in either of them in the same period; and the length of the equivalent simple pendulum is the sum of their radii.

From Equation 60 and the relation $r = R_1 + R_2$ we obtain

$$R_2 = \frac{k^2}{R_1}, \quad R_1 = \frac{k^2}{R_2}, \tag{61}$$

relations which may be used to compute one of the values of R if the other is given.

This property of the pendulum makes it possible to use it to determine the force of gravity without calculation of the moment of inertia. This is done by swinging the pendulum about two points which lie on a line passing through the center of mass, and adjusting the distribution of its parts until the periods of oscillation about these points are equal. If the points are not equidistant from the center of mass, it follows from the theorem just proved that the distance between the points of suspension in the two positions is the sum of the two radii R_1 and R_2 and so is equal to r.

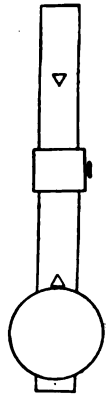


Fig. 84.

The pendulums commonly used in this operation are long rods or bars fitted with fixed knife-edges facing each other at the distance apart which is to be r, and which can be carefully measured (Fig. 84). A heavy weight is attached at one end to insure that the center of mass will not lie midway between the knife-edges; and a mass is also arranged to be moved up and down the bar and fixed in any desired position. The pendulum is swung about both knife-edges in turn and the sliding weight adjusted until the periods of oscillation are equal. When

this condition is attained, the time of oscillation is that of a simple pendulum whose length is the distance between the knife-edges, and by using it in the formula $t = \pi \sqrt{\frac{r}{g}}$, the value of g may be directly calculated.

This pendulum is generally called the reversible pendulum. It was used in a geodetic survey by Kater, and is also often called Kater's pendulum.

By adjusting the fixed parts of the pendulum in the making, the periods of oscillation about the knife-edges can be made nearly equal. If this is done the value of g can be calculated from the observed times of oscillation about the two knife-edges by a formula obtained by a slight extension of the method we have used, and the troublesome adjustment of the sliding mass may be avoided.

107. Center of Percussion. — If an impulse is applied to a point in a free body, the initial motion of the body is a rotation about an axis which is perpendicular to the plane containing the center of mass and the line in which the impulse is applied (Fig. 85). The point at which the line drawn from this axis through the center of mass meets the line of the impulse is called the *center of percussion* for the particular axis around which the initial rotation takes place.

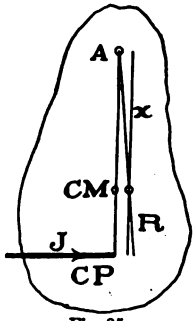


Fig. 85.

To find the distance between the center of percussion and the axis we resolve the impulse (§ 103) into an equal impulse applied to the center of mass and an impulsive couple. The effect of the impulse J acting at the center of mass is to give to the body a motion of translation with the velocity v and the momentum $Mv = J$. The couple has a moment about the center of mass equal to JR, if R represents the distance from the center of mass to the line of the impulse. This couple will give the body an angular velocity ω about the center of mass, and an angular momentum

$$K\omega = JR.$$

As the body moves forward and rotates, some one point in the body, extended if necessary by a massless sheet, will be at rest; because while moving forward with the translation with the velocity v it is moving backward with the rotation with an equal velocity. This point must lie on the line produced which is drawn from the center of mass perpendicularly to the line of the impulse, for it is only in this line that the translation and the rotation can be exactly opposite to each other. Using x to represent the distance of this point from the center of mass so that its velocity due to the rotation is ωx, we have

$$v - \omega x = 0, \quad \text{or} \quad x = \frac{v}{\omega}.$$

Substituting the values of v and ω from the previous equations we find

$$x = \frac{K}{MR} = \frac{k^2}{R}.$$

This gives the distance of the instantaneous axis of rotation from the center of mass. To get the distance between this axis and the line of the impulse, we must add R to x. Writing $x + R = r$, we have

$$r = \frac{K + MR^2}{MR} = \frac{k^2 + R^2}{R}.$$

The formula is the same as that which gives the distance of the center of oscillation of a pendulum from its axis of suspension. The center of oscillation is therefore the center of percussion for the pendulum, and an impulse applied perpendicularly to the line drawn from the axis through the center of mass, will start the pendulum in rotation without bringing strain on the axis of suspension. Similarly any body mounted so as to turn about any axis may be set in rotation without bringing a strain on the axis, by an impulse applied in the proper direction to the center of percussion.

EXAMPLES, IX

1. *To prove the equation of motion of a rotating body by direct study of the angular acceleration.*

Let a force F (Fig. 86) act on a mass m, which is borne on the end of a massless rod, so that it can only move in a circle about a center O. The component of the force along the rod will have no effect on the motion of the mass, but the component at right angles to the rod will give to the mass an acceleration a, so that $F \cos \phi = ma$. Now we may express a in terms of the angular acceleration α and the radius r, for $a = \alpha r$; so that $F \cos \phi = mr\alpha$. If we also substitute $\frac{p}{r}$ for $\cos \phi$, we have finally $Fp = mr^2\alpha$. Now

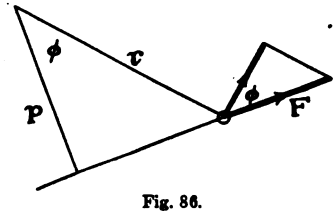


Fig. 86.

$mr^2 = I$, the moment of inertia of the mass m about the axis through O, so that we obtain for a single particle the equation obtained otherwise in § 100.

For a collection of particles rigidly bound together or for an extended body to which any number of forces are applied in planes perpendicular to the axis of rotation we shall obtain the more general formula of § 100. Each particle of the system or body will be acted on by a resultant force to which the formula just found will apply, and the moment of each resultant force is the sum of the moments of its component forces (§ 44). If we therefore write all the equations for all the particles and add, since the angular acceleration must be the same for all the particles of a rigid body, we shall have $\Sigma Fp = \alpha \Sigma mr^2$. The sum of the moments as described contains the moments of the internal forces due to the reactions which arise from the rigid binding of the particles, as well as those of the external forces, but the internal forces always occur in pairs of equal and opposite forces acting in the same line and therefore having equal and opposite moments about any center. The moments of the internal forces therefore drop out of the sum of the moments, and we have finally $\Sigma Fp = I\alpha$, as already obtained in § 100, in which the forces are external forces only.

2. **Acceleration of Sphere Rolling down Inclined Plane.** — In discussing the experiment by which Galileo confirmed the principle that a constant force will produce a constant acceleration, it was remarked that the discussion needed to be supplemented by a proof that the relation between force and acceleration

was not altered by the circumstance that the ball rolled down the plane instead of sliding. As we can now see, some of the work done by the force was employed in giving the ball energy of rotation. It has to be shown that the fundamental proportionality between the force and the acceleration of the ball is not affected thereby.

We call the radius of the ball r and its mass M. The angle of inclination of the inclined plane down which it rolls is ϕ. If the ball rolls without slipping, its point of contact with the plane is instantaneously at rest, and the ball has an angular acceleration α about the point of contact as center. This is given to it by the component of its weight which is acting down the plane and is applied at its center of gravity. This force is $Mg \sin \phi$; its moment about the point of contact is $Mgr \sin \phi$; and the angular acceleration which it will impart to the ball is $\alpha = \frac{Mgr \sin \phi}{I}$. The quantity I is the moment of inertia of the ball about an axis passing through the point of contact, and therefore at the distance r from a parallel axis passing through the center of mass of the ball. The moment of inertia of the ball about this axis is therefore (§ 104) $I = Mr^2 + \frac{2}{5}Mr^2 = \frac{7}{5}Mr^2$. The angular acceleration α equals the linear acceleration a of the center of mass divided by r, or $\alpha = \frac{a}{r}$. Substituting these values for I and α we obtain

$$a = \tfrac{5}{7} g \sin \phi.$$

This acceleration a is that which is determined from the observations, and Galileo's experiments showed it to be constant for any given inclination of the plane. The formula just obtained shows that if a is constant, g also is constant, so that the principal conclusion which Galileo drew from this experiment is justified. If the observations are used to obtain the value of g, we cannot take for it $\frac{a}{\sin \phi}$, as could be done if the ball slid down without rolling; but must multiply $\frac{a}{\sin \phi}$ by the numerical factor $\frac{7}{5}$.

3. **The Atwood's Machine.** — When the Atwood's machine is used to illustrate the effects of a constant force, the overweight not only sets the masses on the cord in linear motion, but also sets in rotation the pulley over which the cord passes. It is necessary to investigate this case also, to see that the fundamental relations illustrated by the Atwood's machine are not affected by this circumstance.

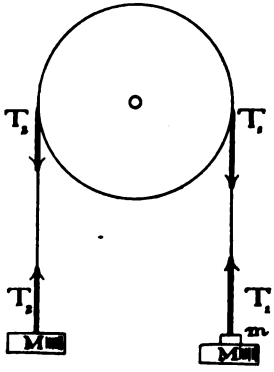


Fig. 87.

Let us represent each of the two equal masses hung on the cord by M and the mass of the overweight by m; the moment of inertia of the wheel by I and its radius by R. The mass $M + m$ on the overweighted end of the cord (Fig. 87) is acted on by its weight $(M + m)\,g$ directed downwards and by a tension T_1 in the cord directed upwards. Its acceleration a will therefore be given by the equation

$$(M + m)\,a = (M + m)\,g - T_1.$$

Similarly the same acceleration a will be given in terms of the tension T_2 acting on the other mass by

$$Ma = T_2 - Mg.$$

The two tensions act on the rim of the wheel to turn it in opposite senses, and the angular acceleration α of the wheel is obtained from the equation

$$I\alpha = (T_1 - T_2)\, R.$$

Now if the cord turns the wheel without slipping on the rim, we have the acceleration a and the angular acceleration α connected by $a = \alpha R$, and with this value for α the last equation becomes

$$Ia = (T_1 - T_2)\, R^2.$$

The tensions are unknown, but they can be eliminated from this equation by the help of the two others in which they appear. Carrying out the substitutions and transposing to obtain the value of a we get

$$a = \frac{mg}{(2M + m) + \dfrac{I}{R^2}}.$$

The experiments show that a is constant, and the formula shows that if that is so, g also is constant, so that the principal conclusion which was drawn from the Atwood's experiment was justified. If the observations are used to obtain the value of g, the moment of inertia of the wheel must be found and taken into account.

4. *To find the value of g with a rigid pendulum with two knife-edges.*

Let t_1 and t_2 represent the times of oscillation about the two knife-edges; let h_1 and h_2 represent the distances from the knife-edges to the center of gravity of the pendulum. Then we have

$$t_1 = \pi\sqrt{\frac{k^2 + h_1^2}{gh_1}}; \quad t_2 = \pi\sqrt{\frac{k^2 + h_2^2}{gh_2}};$$

whence

$$\frac{\pi^2}{g} = \frac{t_1^2 h_1}{k^2 + h_1^2} = \frac{t_2^2 h_2}{k^2 + h_2^2} = \frac{t_1^2 h_1 - t_2^2 h_2}{h_1^2 - h_2^2}.$$

This can be broken up into partial fractions, and we have

$$\frac{\pi^2}{g} = \frac{t_1^2 + t_2^2}{2(h_1 + h_2)} + \frac{t_1^2 - t_2^2}{2(h_1 - h_2)}.$$

The sum $h_1 + h_2$ is the distance between the knife-edges, if they are so placed that the center of gravity is on the line joining them, and can be accurately measured. The distances h_1 and h_2 cannot be so accurately measured, but if the periods t_1 and t_2 are nearly equal, the second term is small, and any slight inaccuracy in determining h_1 and h_2 will not affect the value of g. When $t_1 = t_2$, the formula reduces to the simple one already obtained for the reversible pendulum.

5. *Where should a billiard ball be struck so that it will roll without sliding?*

The axis of rotation is the lowest point of the ball. Call the radius of the ball R. The distance r of the center of percussion from the lowest point is $r = \dfrac{k^2 + R^2}{R}$, and $k^2 = \frac{2}{5}R^2$ for a sphere. Hence $r = \frac{7}{5}R$.

6. *If a ball is released at some point near the bottom of a spherical bowl, in which it rolls, to find its motion, neglecting the retarding effect of friction.*

The ball will roll to the lowest point of the bowl, and after passing through it,

will ascend on the other side. The path of its center will be a small arc of a vertical circle whose radius r is the distance from the center of the ball to the center of the sphere of which the bowl forms a part.

The moving force at any point at which the tangent to the surface makes the angle ϕ with the horizontal is $Mg \sin \phi$. The moment of this about the point of contact of the ball with the surface is $MgR \sin \phi$, and the angular acceleration of the ball about that point is $\alpha = -\frac{MgR \sin \phi}{I}$. The moment of inertia I (see Example 2) is $\frac{7}{5}MR^2$; the tangential acceleration a of the ball is αR. Making the substitutions we have

$$a = -\tfrac{5}{7}g \sin \phi.$$

Now ϕ is small, so that $\sin \phi = \phi$, and if we represent the distance of the center of the ball from the lowest point of its path by x, we have $\phi = \frac{x}{r}$, and

$$a = -\frac{5}{7}\frac{g}{r}x.$$

The acceleration is therefore proportional to the displacement, and the motion of the ball is simple harmonic. The period is given by the formula (§ 63)

$$t = \pi\sqrt{\frac{7r}{5g}}.$$

General Motion of Rigid Bodies

108. Motion in Three Dimensions. — The general motion of a rigid body, and even the apparently simple case of motion of a body around a fixed point, whether acted on by forces or not, is of great complexity, and a discussion of it cannot be undertaken in this book. Elementary methods, such as we have been able to use up to this time, are not applicable even to the most simple problems. Correct solutions can be obtained only by the use of the methods of advanced dynamics. The consideration which we shall give to a few of the most intelligible and interesting facts connected with the subject will be by description only.

The motion of a body in three dimensions may be described in a way which is in some respects analogous to that employed for the description of the motion of a plane figure. Any displacement of a body may be accomplished by a translation and a rotation around a suitably chosen axis. A direction of translation may always be found such that the axis around which the necessary rotation takes place is in the same direction. An infinitesimal displacement of the body may thus be analyzed into an infinitesimal translation and an infinitesimal rotation around an axis drawn in the same direction as the translation. This motion is that of a part of a screw, when the screw is turned and so driven forward, and the motion of a

body at any instant may therefore be called a screw motion. As the body moves through a finite path, the characteristics of the screw whose motion describes the motion of the body, that is, the pitch of the screw and the direction of its axis, change from instant to instant. The changes, however, are never discontinuous.

109. Principal Axes of Inertia. — The study of the moment of inertia of a body around an axis passing through the center of mass shows that this moment of inertia can always be found if we know the moments of inertia around three principal axes which pass through the center of mass and are perpendicular to each other. Around one of these axes the moment of inertia has a maximum value, around one of the others a minimum value. When the body is rotating around its axis of greatest moment of inertia, or around its axis of least moment of inertia, it is in a condition of kinetic stability, that is, an impulse applied to the body, though it will compel the body to rotate around a new axis, will not so alter the motion of the body as to cause the new axis to deviate more and more from the direction of the original axis. If the original rotation is around the third of the principal axes, its condition is unstable, that is, though the rotation will persist around that axis so long as no impulse is applied to the body, the application of an impulse will cause a rotation around a new axis, whose direction will continually deviate more and more from that of the original axis. A rotation set up around any other axis than one of these three will not continue around that axis, even though no impulse is applied to the body, but the axis around which rotation occurs will change its direction in the body continually.

110. Motion of a Body with One Point Fixed. — When a body has one point fixed, its only motion is rotation about that point. The case of special interest is that in which the body is a figure of revolution about an axis, one point of which is fixed, and is given a great initial angular velocity about that axis. The top spinning on a rough floor, so that the point of its peg is fixed, is an example of such a body. The common gyroscope is another example.

The gyroscope is a heavy wheel suspended by the terminal points of its axis in two gimbal rings, one of which can turn about a horizontal axis perpendicular to that of the wheel, the other about a vertical axis. The three axes are mutually perpendicular, and the wheel has complete freedom of rotation about its middle point.

When the wheel of the gyroscope is rapidly spun, it possesses kinetic stability, so that an impulsive couple applied to it will set it oscillating about its original position. Speaking generally this is due to the fact that when the body is thrown out of the state of steady rotation about its axis of figure, centrifugal couples are developed which react against the impulse, and change their position in the body as the axis of rotation changes.

When the wheel is spun with its axle horizontal or nearly so, and a weight is hung on the inner gimbal ring, so as to introduce a couple tending to rotate the wheel around a horizontal axis perpendicular to the axle, the first effect is generally to cause an initial slight rotation around this axis; but as soon as this begins, the centrifugal couples developed by the change in the axis of rotation come into play to cause a rotation of the body around the vertical axis, and accompanying this, a rotation in the opposite sense to that introduced by the weight around the horizontal axis. As the weight falls the rotation around the vertical axis increases, and with it the accompanying rotation around the horizontal axis; so that at last the fall of the weight is checked, and the weight is lifted, while the rotation around the vertical axis diminishes. As the weight again falls, the rotation around the vertical axis again increases, and the weight is again lifted; and so on. The theory of this motion shows that the original rotation of the wheel around its axle will remain unchanged, and that the motions of the end of the axle introduced by the weight may be resolved into two, one of which, called the *precession*, is a uniform motion in a horizontal circle, and the other, called the *nutation*, is a periodic motion in a small ellipse. The velocity of precession is inversely as the angular velocity of the wheel around its axis, and is generally slow. The period of the nutation is inversely as the angular velocity of the wheel, and is small when the wheel is spinning rapidly.

Since the nutation arises from the changing vertical and horizontal velocities of the ends of the axis, it may be avoided if an initial angular velocity is given the wheel about the vertical axis, equal to the precession which the weight would cause. The centrifugal couples introduced by this rotation will exactly counteract the effect of the weight, so that the only motion in this case will be a regular precession, without nutation.

The motion of a top, if its initial axis of rotation is sufficiently inclined to the vertical, exhibits precession and nutations like those of the gyroscope. When the initial axis of rotation is nearly vertical, the axis performs small oscillations

about the vertical, the end of the axis tracing out some form of epicycloid. The axis of the top does not depart far from the vertical and its condition is stable. In the ordinary case, in which the peg is somewhat blunt, so that the point of contact with the ground, as the top is inclined, is not in the axis of rotation, the friction between the revolving peg and the ground introduces another couple, of which the effect is to tilt the axis upward; so that the end of the axis will rise and finally become vertical.

These examples illustrate the general principle that steady rotation of a body about an axis introduces a resistance to any force which is so applied as to change the direction of that axis. Thus if a heavy wheel is mounted inside a box and is kept in continual rotation, the box, though of itself it does not originate motion, or resist a motion of translation, will offer a resistance to any force tending to turn it around. Certain forces which exist in nature have been explained by supposing that the bodies which exhibit them contain portions which are in rapid rotation. Since these rotating portions, even if they exist, are such that they can never be perceived, they are said to be concealed, and the forces exhibited by the bodies are said to be due to concealed motion.

111. **Composition of Angular Velocities.** — We may represent an angular velocity in a diagram by drawing a line from some origin to represent the axis of rotation, in such a direction that the rotation will appear counterclockwise to one looking along it from the origin, and measuring off on that line a length equal to the angular velocity. An angular displacement may also be represented by a similar construction. Magnitudes which may be represented in this way are vectors of a certain sort, which have been called *localized vectors*, or *rotors*. They cannot be transferred from place to place in a construction as ordinary vectors can be, and the rules for their composition are in general different from those for ordinary vectors. In any case the rules for dealing with them must be derived by independent study of their properties, and cannot be inferred from the general rules relating to ordinary vectors.

When we consider two finite rotations about independent axes, we find that the displacement of a body to which they are applied will be different according to the order in which they are applied. When the rotations are infinitesimal, the order of application becomes indifferent; and hence it becomes possible to give simple rules for the composition of angular velocities. We shall consider only the case of angular velocities about axes which pass through a common point. Two angular velocities about axes meeting at a

point are equivalent in their effect upon a body to which they are applied to a single angular velocity about an axis passing through the same point, and found by the following construction: The resultant angular velocity is represented by the diagonal of a parallelogram of which the sides are the lines representing the component angular velocities.

To show this we notice that the velocity of any point, whose distance from the axis of rotation is p, is ωp, if ω is the angular velocity. Consider two intersecting axes (Fig. 88) about which the angular velocities are ω_1 and ω_2. The magnitudes ω_1 and ω_2 are represented by the lengths of the lines $\overline{OA}$, $\overline{OB}$. As the rotations are both counterclockwise as seen from O, the rotation around $\overline{OA}$ will carry the body upward through the paper in the angle between the axes, and that around $\overline{OB}$ will carry it downward through the paper in the same angle. There will therefore be some point, say P, at which the two velocities imparted by the two rotations will counteract each other, so that we shall have

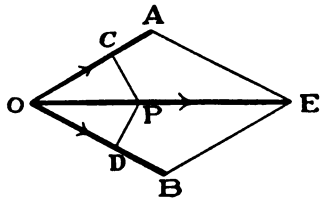


Fig. 88.

$$\overline{PC}\,\omega_1 - \overline{PD}\,\omega_2 = 0.$$

Now $$\overline{PC} = \overline{OP}\sin POA, \qquad \overline{PD} = \overline{OP}\sin POB,$$

so that the condition which determines the point P, at which there is no motion, is

$$\omega_1 \sin POA = \omega_2 \sin POB.$$

But this is the condition that the point P lies on the diagonal of the parallelogram of which ω_1 and ω_2 are the lengths of the sides. This line of no motion may therefore serve as an axis of rotation.

To find whether an angular velocity about OP as axis will give the same motions to all points of the body that are given to them by the angular velocities about $\overline{OA}$ and $\overline{OB}$, we select any point S (Fig. 89), and determine its motion. If we designate by p_1 and p_2 the perpendiculars from S on the axes $\overline{OA}$ and $\overline{OB}$, we may write the velocity v of the point S as

$$v = p_1\omega_1 + p_2\omega_2.$$

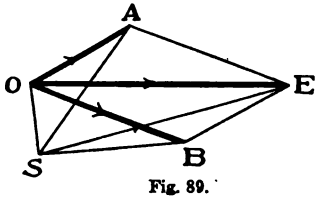


Fig. 89.

Now since ω_1 is the base of the triangle OSA, and since p_1 is its altitude, the product $p_1\omega_1$ is equal to twice the area of the triangle OSA; and similarly $p_2\omega_2$ is equal to twice the area of the triangle OSB. If the parallelogram $OAEB$ is constructed on $\overline{OA}$ and $\overline{OB}$, and the point S joined to the point E, the triangle OSE stands on a common base OS with the triangles OSA and OSB, and has an altitude equal to the sum of their altitudes, from the geometry of the parallelogram. Its area is therefore equal to the sum of their areas, so that we have

$$\text{Area of } OSA + \text{Area of } OSB = \text{Area of } OSE.$$

But by our previous proof OE is the line of no motion, and may be taken as an axis of rotation; and if we assign to the rotation around it an angular velocity equal to ω, represented by the length OE, we may represent the velocity of the point S due to this rotation by $p\omega$, if p represents the perpendicular on OE from S. As before $p\omega$ equals twice the area of the triangle OSE, and from the equation among the areas we get also

$$v = p_1\omega_1 + p_2\omega_2 = p\omega;$$

so that the velocity of any point S may be represented as resulting from rotation about the axis OE with the angular velocity ω, represented by $\overline{OE}$, the diagonal of the parallelogram of which $\overline{OA} = \omega_1$ and $\overline{OB} = \omega_2$ are sides.

The point S has been taken in the plane of the axes. The proof required to show that the parallelogram construction will give the true angular velocity of any point of the body, whether in the plane of the axes or not, is too elaborate for insertion here.

112. Foucault's Pendulum. — The use of this construction in dealing with rotations may be illustrated by applying it to the description of the behavior of the Foucault pendulum. This pendulum consists of a heavy bob suspended by a long cylindrical wire. The upper end of the wire is held in a clamp in such a way that it may swing with equal freedom in any direction. The pendulum thus arranged is drawn aside and released as carefully as possible, so as to give it no sideways swing, and the points which it reaches at the ends of its swings are observed from time to time. It is found that the path of the bob is a very eccentric ellipse, and that the major axis of this ellipse, which marks what we may call the plane of swing, changes its apparent direction with respect to the surface of the earth. The angular deviation of this plane of swing from its original position, in a given time, is at different places proportional to the sine of the latitude of the place.

We may give a kinematic description of this result as follows: If we conceive such a pendulum set up at the North Pole, it will swing in a plane which will remain fixed in space and the earth will turn around under it, so that to an observer examining the pendulum, its plane of swing will change its position relative to the earth's surface. In one day it will appear to have traversed a complete circle. If the same pendulum is set up at the Equator, and is set swinging in the north and south line, the earth will carry it around, and its plane of swing will not change its position relative to the earth. At any intermediate station the motion of the plane of swing may be determined as follows (Fig. 90): Draw a line from the center of the earth to the station at P and another line OD perpendicular to this, in the plane containing this line and the

earth's axis. We then apply the theorem of the last section by resolving the angular velocity of the earth into two angular velocities around these two axes. The angular velocity around OD will have no effect upon the motion of the pendulum. The angular velocity about the other axis OP, however, which is equal to the angular velocity of the earth multiplied by the sine of the latitude, may be considered as an angular velocity with which the earth turns under the pendulum around the axis passing through the station. The plane of swing will therefore have an apparent angular velocity around this axis which is proportional to the sine of the latitude. The dynamical discussion of the motion of the Foucault pendulum cannot be given here.

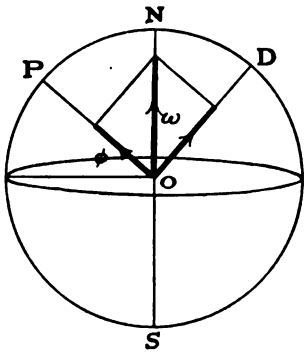


Fig. 90.

Kinetic Friction

113. Friction. — When a solid body is moved over the surface of another solid body, as when a grindstone turns under an ax blade, or when a body moves through a fluid, as when a ball rises or falls through air, or a boat moves through water, its motion is always opposed by a force oppositely directed to the motion. This force is called the *force of friction*, or simply *friction*.

Friction occurs so universally in the ordinary cases of motion which are treated in mechanics that it is important to introduce a brief discussion of its purely mechanical relations in this connection.

114. Coefficient of Sliding Friction. — When one solid slides over another, experiment shows that the effect of friction may be represented approximately by adding to the system of forces acting on the body another force directed oppositely to the motion, and therefore tangential to the surface over which the body is sliding. This force is proportional to the force acting upon the body normally to the surface of contact, and depends upon the nature of the two surfaces in contact. It is independent of the velocity with which the body moves.

If the normal force with which the moving body is pressed against the surface is W, and if P represents the force which acts on it

parallel to the surface to move it, and R the reaction of the surface, including the force due to friction, the resultant of these forces is a force F parallel to the surface and in the direction of motion. If the mass of the body is m and its acceleration a, we shall have $F = ma$. The force F is obtained by compounding the various forces (Fig. 91). The reaction R may be resolved into two components, R_1 and R_2, normal to the surface and parallel to it. The reaction R_1 is equal to the normal force W. The force F that moves the body is then equal to the difference between P and R_2; so that we have

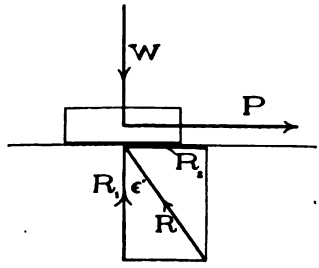


Fig. 91.

$$P - R_2 = ma.$$

If P is greater than R_2 the body will have a positive acceleration; if less than R_2, it will have a negative acceleration, and the body will move slower and slower until it comes to rest. If P is equal to R_2 the acceleration will be zero, and the velocity of the body will be constant.

Since we know by experiment that R_2 is proportional to W, we may write

$$R_2 = \mu' W. \tag{62}$$

The factor μ' is called the *coefficient of sliding friction.* This formula is similar to the one which defines the coefficient of static friction (§ 47). The coefficient μ in that formula is generally greater than the coefficient μ' for the same pair of surfaces; that is, it requires a greater tangential force to start a body sliding over another than it does to keep it moving after it has started.

115. **The Angle of Sliding Friction.** — If we determine the force P' which is just equal to R_2, so that the body under its action slides with a constant velocity, we may find the value of μ' from the formula $\mu' = \dfrac{P'}{W}$.

From Fig. 91 we see that if ϵ' represents the angle between the normal to the surface and the total reaction R, we have $\tan \epsilon' = \dfrac{R_2}{R_1} = \dfrac{P'}{W}$, so that $\mu' = \tan \epsilon'$.

The angle ϵ' is called the *angle of sliding friction.*

In ideal problems bodies sliding over each other are often supposed to be perfectly smooth or frictionless. In such a case the

coefficient of friction vanishes, and the angle of friction becomes zero, so that the reaction of a smooth surface on a body is normal to the surface. In actual arrangements this condition is never attained, but highly polished bodies will sometimes show a very small coefficient of friction.

116. Fluid Friction. — The laws of the friction of a fluid on a solid moving through it are not so simple as the laws of sliding friction. Without going into the question further at this time, it may be said in general that the frictional reaction corresponding to the R_2 of the previous paragraphs is a function of the velocity and increases as the velocity increases. In case the velocity of a body in different fluids is the same, the friction experienced by it will depend on the nature of the fluids, and the acting forces required to maintain the constant velocity will be proportional to the coefficients of friction of the various fluids.

EXAMPLES, X

1. *A body is placed on an inclined plane, and the plane is tilted until the body, when started moving, slides down it with a constant velocity; use this experiment to determine the coefficient of sliding friction.*

Let m represent the mass of the body, and α the angle of inclination of the plane. The moving force P' is $mg \sin \alpha$, the normal force W is $mg \cos \alpha$. We have therefore $\mu' = \frac{P'}{W} = \tan \alpha$; and the α thus found is the angle of friction ϵ'.

2. *Find the weight which must be hung on the end of a cord passing over a pulley at the end of a table, to give to a mass M_1 fastened to the other end of the cord and resting on the table, an acceleration a, the coefficient of sliding friction being μ'.*

The two masses have the same acceleration. Let T represent the tension in the cord. The equation of motion of the unknown hanging weight, whose mass we may call M_2, is $M_2 a = M_2 g - T$.

The reaction due to friction on the mass M_1 is $\mu' M_1 g$, and hence the equation of motion of the sliding weight is $M_1 a = T - \mu' M_1 g$.

Eliminating T we obtain for M_2

$$M_2 = M_1 \frac{\mu' g + a}{g - a}.$$

If M_2 is given instead of a, we obtain, for the acceleration, $a = \frac{M_2 - \mu' M_1}{M_2 + M_1} g$.

3. *If a sled of mass M is set moving over level ice with the velocity v, how long will it move before it comes to rest, if the coefficient of friction is μ, and what distance will it traverse?*

The force resisting the motion is μMg. The acceleration $a = \frac{v}{t}$ is obtained from $Ma = \mu Mg$, and hence $t = \frac{v}{\mu g}$. From the equality between the kinetic energy and the work done by the friction we have $\frac{1}{2} M v^2 = \mu Mgs$, and hence $s = \frac{v^2}{2\mu g}$. The time and distance are independent of the mass.

4. *A long lever is balanced across a horizontal shaft, and is so arranged with a clutch by which it can be clamped to the shaft that as the shaft turns and tends to carry the lever around with it, a weight hung on the lever will prevent its turning, so that the shaft revolves in the clutch and the lever remains at rest and horizontal. Find the work that is being done in turning the shaft.*

The moment of force due to the friction is equal to the moment of force due to the weight. Representing by M the force due to friction and by r the radius of the shaft, by P the weight in pounds and by R its distance in feet from the center of the shaft, this equality gives $Mr = PR$. The shaft makes n turns per minute and so has an angular velocity of $2\pi n$ per minute. The work done per minute in turning the shaft against the friction of the clutch is therefore $2\pi nMr = 2\pi nPR$ foot pounds per minute. The Prony brake or Prony dynamometer, used for measuring the power transmitted by shafting, is essentially the arrangement here described.

5. *If a grindstone of mass M and radius r is rotating n times a second, and an ax is pressed against it with a force of P pounds weight, how long will it turn before coming to rest; the coefficient of friction being μ, and no force being applied to the crank?*

The moment of inertia of a cylinder about its central axis is $\frac{Mr^2}{2}$. The negative angular acceleration due to the moment of force set up by the friction is $\alpha = \frac{\mu Pr}{\frac{Mr^2}{2}} = \frac{2\mu P}{Mr}$. The initial angular velocity is $\omega = 2\pi n$. This is reduced to zero in the time $t = \frac{\omega}{\alpha} = \frac{\pi nMr}{\mu P}$.

6. *A chain is dragged at a uniform rate off a rough road on to a sidewalk: its coefficients of friction are with the road μ_2, with the sidewalk μ_1. Find how the force used to drag it changes with the time.*

Call the length of the whole chain l; of that part on the sidewalk s; and represent by p the weight of unit length of the chain. Then the force used will be $F = \mu_1 ps + \mu_2 p\,(l - s)$.

When none of the chain is on the sidewalk we have $s = 0$, and the force $F_0 = \mu_2 pl$. In general $F = F_0 - sp\,(\mu_2 - \mu_1)$.

If we use v to represent the velocity of the chain, and t the time which has elapsed from the instant the chain began to touch the sidewalk to the instant at which s is measured, we have $s = vt$, and hence $F = F_0 - t \cdot vp\,(\mu_2 - \mu_1)$.

For a given velocity the force falls off by decrements which are proportional to the time.

When t becomes so great that $vt = l$, at which time the chain is all on the sidewalk, we have $F = \mu_1 lp$. This value for it remains thereafter constant.

CHAPTER IV

MECHANICS OF LIQUIDS

HYDROSTATICS

117. The Problem of the Crown. — The study of the peculiar mechanical effects exhibited by liquids was begun by Archimedes. According to the story told by Vitruvius, a certain quantity of gold had been given by King Hiero to a goldsmith, with the order to use it in constructing an elaborate crown. The crown when returned weighed as much as the gold which had been supplied, but Hiero suspected that a part of the gold had been abstracted, and silver substituted for it. He asked Archimedes to devise a way to determine whether this was so or not, without injuring the workmanship of the crown. Archimedes is said to have discovered how this might be done by noticing the way in which the water overflowed from a bath into which he had entered. This observation suggests to any one who considers it that, since bodies of equal weights do not generally occupy equal volumes, when the materials of which they are constituted are different, the one having the larger volume will cause more water to overflow from a full vessel in which it is immersed than the other one will, and that by determining the overflow caused by the crown and by equal weights of gold and silver, the question asked of Archimedes might be answered.

By reflection upon the subject thus brought to his attention, Archimedes was led to study the properties of liquids at rest or in equilibrium. He presented his conclusions as a series of deductions from the postulates of his theory, these postulates being essentially a definition of a liquid. According to Archimedes a liquid is a continuous body of such a nature that when two portions of it are similarly situated and are contiguous to each other, the portion which is under less pressure is set in motion by the portion which is under greater pressure; and that the pressure at a point in it under a free surface is proportional to the depth of the point below the surface. From these postulates Archimedes was able to deduce, by purely formal reasoning, the laws which govern the apparent loss of weight of bodies immersed in a liquid, and to determine certain cases of equilibrium of floating bodies.

Archimedes' method, while perfectly correct, does not give so clear a presentation of the fundamental notion of pressure as

that afterwards adopted by Pascal, and we shall accordingly pass, without further discussion of it, to the consideration of Pascal's treatment of the equilibrium of liquids.

118. **Pascal's Fundamental Experiment.** — Pascal does not give a formal definition of a liquid, but assumes it to be continuous, incompressible, and fluid. By its being fluid we may understand that its parts are not held together by any internal forces which prevent their sliding past each other, so long as such a motion does not change the density of the liquid. When a force is applied which will cause such sliding of one part over another, the liquid yields to it, and equilibrium is not attained until such forces have vanished.

Pascal's method is, ostensibly at least, based on a fundamental experiment. He describes a closed vessel (Fig. 92) furnished with two vertical tubes of different sizes, in which slide closely fitting pistons. The vessel is entirely filled with water. He then says that if a force, such as a weight, is applied to one of the pistons, it will move in and the other one will move out, unless a force is applied to it also, so adjusted that the two forces are proportional to the areas of the pistons to which they are applied.

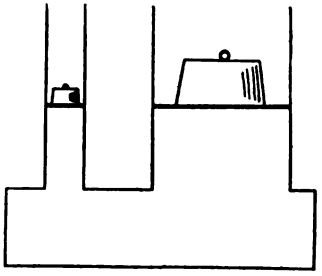


Fig. 92.

This experimental result is further established by being shown to be in harmony with the principle of work. For, as the volume of the water remains constant, when one of the pistons moves in one sense the other moves in the opposite sense, and the distances over which they move are inversely as their areas; and as the forces applied to them are directly as their areas, the products of the forces and the distances through which their points of application move are equal in magnitude and opposite in sign; so that the work done by the forces is equal to zero, and the forces are in equilibrium according to the principle of work. In this way of looking at the experiment the liquid is simply the medium for the transmission of force from one of the points of application to the other, and otherwise does not enter into consideration in determining the conditions of equilibrium.

119. **Hydrostatic Pressure.** — Pascal, however, also looked at the matter in another way. He considered the liquid as exerting force against each piston, and thus holding both forces in equilibrium. A force is exerted by each element of the liquid which is in contact with the surface of the piston, or indeed with any part of the vessel. This force is perpendicular to the surface, for other-

wise the parts of the liquid would slide past each other along the surface, and thus not conform to the condition of equilibrium. The sum of all these elementary forces applied to the surface of one of the pistons is a resultant force perpendicular to the piston and equal to the opposite force applied to it (Fig. 93). Similar elementary forces act within a liquid from one portion of it against another across the surface which separates them.

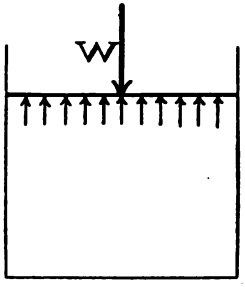


Fig. 93.

In the case of Pascal's experiment the resultant forces on the pistons are proportional to their areas, so that the resultant force applied to each unit of area is the same for both pistons. If we represent this resultant force per unit of area by p, the total force by W, the area by a, we have

$$p = \frac{W}{a}. \tag{63}$$

In this case, in which the forces are uniformly distributed over the whole surface, we say that the pressure exerted on that surface by the liquid is uniform, and if the surface is plane, is measured by the ratio of the force exerted on it to its area. The equation $p = W/a$ then defines and measures the *pressure*.

If the pressure is not uniform, as it never is unless the weight of the liquid can be neglected, the formula just given defines and measures the average pressure over a finite surface of area a, or if the area is reduced until the pressure on it may be considered uniform, or if the limit of the ratio is taken, as the area a approaches zero, the same formula defines and measures the pressure at a point.

Pressures such as we have now considered are similar to the forces by which they are defined in being perpendicular to the surfaces upon which they act. They are the only pressures which can exist in a liquid in equilibrium, for pressures tangential to any surface would set up a flow of the liquid along the surface, and thus not conform to the condition of equilibrium. They are therefore called *hydrostatic pressures*.

120. Units of Pressure. — The unit of pressure in general is a pressure equivalent to a unit force applied to each unit of area. In the c.g.s. system it is the pressure equivalent to the force of one dyne per square centimeter. For use in engineering practice,

this unit is inconveniently small, and instead of it a pressure of one kilogram weight per square centimeter is often used. In English-speaking countries the usual practical unit of pressure is the pressure of one pound weight per square inch.

The pressure of one atmosphere, which is also often used as a unit, will be defined in § 128.

121. Pascal's Principle. — From his fundamental experiment, and his conception of pressure, Pascal deduced what is known as *Pascal's principle:*

In any liquid which is in equilibrium under the action of no forces except those which may be applied to its bounding surface, the pressure at every point, and in every direction about any point, is the same.

This principle may be obtained without experiment from the condition defining a liquid, that the pressure on any surface drawn in it is always perpendicular to the surface. To show this choose a point in the liquid as origin (Fig. 94), draw three mutually perpendicular lines from it, and through three points on those lines, very near the origin, pass a plane. This plane, and the three planes determined by the lines drawn, limit a minute tetrahedron, the liquid in which is in equilibrium under the forces on its four faces. Let us represent the area of the face ABC by S, and the angles made by the normal to it with the three other lines OX, OY, OZ, by α, β, γ respectively. Then the areas of the other faces are

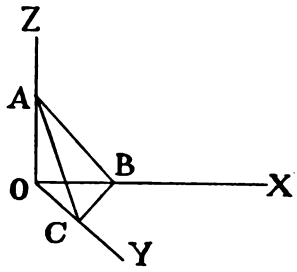


Fig. 94.

$$OAC = S\cos\alpha, \quad OAB = S\cos\beta, \quad OBC = S\cos\gamma.$$

Represent the force applied to the face ABC by R. Its components along the three lines OX, OY, OZ are in equilibrium with the forces X, Y, Z applied to the faces perpendicular to those lines, so that we have

$$X = R\cos\alpha, \quad Y = R\cos\beta, \quad Z = R\cos\gamma.$$

The pressures on the faces are found by dividing these forces by the areas to which they are applied. When this is done, it appears that the pressures are all equal, and given by R/S, the measure of the pressure on the face ABC; and as no restriction was set on the position in which the face ABC was placed, it is thus shown that the pressure at a point in a liquid is the same in all directions.

To prove that the pressure is the same at any two points in the liquid we describe a right cylinder in the liquid around the line joining the two points as axis (Fig. 95). The liquid in this cylinder is in equilibrium under the pressures on its surfaces. The pressures on the various parts of the cylindrical surface are perpendicular to it, and so will have no components parallel with the axis. The liquid is therefore in equilibrium, as regards motion along the axis, because of the

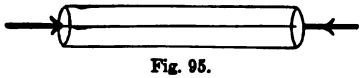


Fig. 95.

equality of the forces acting at the end faces, and since these are equal and parallel to each other, the pressures on them at the given points are equal.

The demonstration that the pressure is the same in all directions at a point holds good even when the pressure in the liquid varies from point to point because of the weight of the liquid, for the volume of the tetrahedron, and so also its weight, is a quantity of the third order of small quantities, while the areas of its faces, and so the forces applied to them, are quantities of the second order, so that the weight drops out of the equation of equilibrium. On the other hand, the proof that the pressure is the same at all points does not hold good when the weight of the liquid is taken into account.

122. Pressure in a Liquid Due to Its Weight. — When a liquid stands at rest in an open vessel, the pressure at any point within it, due to its weight, is proportional to the depth of that point below the surface. Let us imagine a circle of unit area (Fig. 96) placed at the point and parallel with the surface. The cylindrical column of liquid which stands on that area is in equilibrium under the combined action of its weight, of the upward pressure applied to the area on which it stands, of the downward pressure applied to the area at the surface, and of the pressures applied to its sides. The pressures applied to its sides are perpendicular to the cylindrical surface and exert equal and opposite forces upon this cylinder, so that they do not counteract the weight of the column. The weight is counteracted only by the excess of the upward pressure applied to the base of the cylinder over the downward pressure applied to the top, and since the weight is proportional to the height of the cylinder, the excess of upward pressure is also proportional to that height.

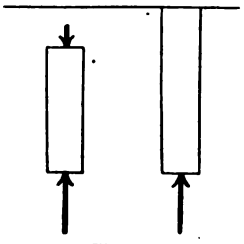


Fig. 96.

123. Free Surface of a Liquid. — The free surface of a liquid is a horizontal plane. If we describe a cylinder (Fig. 97) in the liquid around a horizontal line as axis, the pressures acting on its two ends are equal; for the cylinder is in equilibrium along its axis under these pressures alone, since its weight has no horizontal component. The two points at which the pressures are equal will, from the previous proposition, be at equal distances from the surface, and the surface is therefore horizontal.

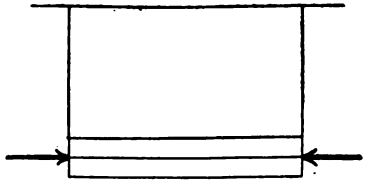


Fig. 97.

If the mass of liquid is so extended that the weights of its different parts are not parallel, but converge toward the center of the earth, a slight modification of the demonstration just given will show that the free surface is part of the surface of a sphere, whose center is the center of the earth. This form of the proposition was demonstrated by Archimedes.

From this proposition it may easily be seen that a liquid in two communicating vessels will stand at the same level in both.

124. Archimedes' Principle. — One of Archimedes' propositions is of such fundamental importance that it is usually recognized as his especial contribution to hydrostatics, and is called *Archimedes' principle.* We may state it as follows:

When a body is immersed in liquid it loses in apparent weight an amount equal to the weight of the liquid displaced by it.

Of course the body does not really lose weight, but part of its weight is counteracted by the upward pressure of the liquid. The truth of this principle may be seen if we imagine the body removed from the liquid and the space which it occupied filled with the liquid. The whole liquid mass will then be in equilibrium, and the weight of the portion which has been introduced will be equal and opposite to the resultant of the pressures applied to its surface by the surrounding liquid. Exactly these same pressures, having the same resultant, were applied to the body when it was in the liquid, and its apparent weight is therefore less than its real weight by an amount equal to this resultant force, or to the weight of the liquid which it displaces

The method which Archimedes is said to have used in the actual solution of the problem of the crown will illustrate this principle. The apparatus (Fig. 98) which he employed was an equal armed lever or balance, on one arm of which a sliding weight could be moved. A weight of gold equal to the weight of the crown was hung on one end of this balance and an equal weight of silver on the other. The masses of gold and silver were next immersed in water. The balance was then no longer in equilibrium, the gold appearing to be heavier than the silver. Equilibrium was restored by placing the sliding weight at a certain distance from the point of suspension. It is plain, from Archimedes' principle, that the moment of force introduced by this weight was equal to the moment of force in the opposite sense introduced by the excess of upward pressure exerted by the water on the silver over that exerted on the gold. This moment, therefore, was proportional to the excess of volume of the silver over that of the gold. The crown was next substituted for the silver, and the gold and the crown immersed in water. Since the crown contained silver, its volume was greater than that of the gold, and there was an excess of

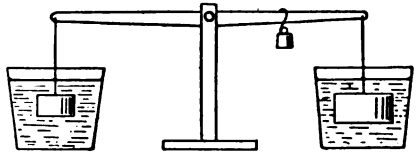


Fig. 98.

upward pressure on it. Equilibrium was restored by placing the sliding weight at another distance from the axis of suspension. By reasoning similar to that already employed, we conclude that the moment of force introduced by the weight in this case was proportional to the excess of the volume of the crown over that of the gold. This excess was due to the silver in the crown, and proportional to the amount of silver present, so that the ratio of the quantity of silver in the crown to the quantity of silver which weighed as much as the crown, was equal to the ratio of the moments of force observed in the two experiments, or to the ratio of the arms to which the sliding weight was applied.

125. Floating Bodies. — When a body is so large and weighs so little that it displaces, when completely submerged, a weight of liquid greater than its own weight, it will not sink in the liquid. It will only be immersed so far that the weight of the liquid which it displaces is exactly equal to its own weight. It then floats upon the liquid.

When such a body is placed upon the surface of the liquid it will generally not remain at rest, but will turn round until it attains some definite position of equilibrium. If it is afterwards slightly disturbed from this position, it will come back to it again if its equilibrium is stable. When it is in equilibrium (Fig. 99), its weight W applied at the center of gravity C and the resultant of the upward pressures which are applied to its surface must be equal and opposed to each other in the same vertical line. The center of gravity D of the liquid which is displaced by the body is the point at which the resultant P of the upward pressures may be considered as applied. The vertical line joining the center of gravity of the body and the center of gravity of the displaced liquid when the body is in equilibrium is called the *axis* of the body. If the axis is tilted to any degree, the shape of the submerged part of the body generally changes, and the center of gravity of the displaced liquid is displaced to one side of its former place. The resultant upward pressure and the weight of the body then form a couple (Fig. 100) which turns the body back toward its original position.

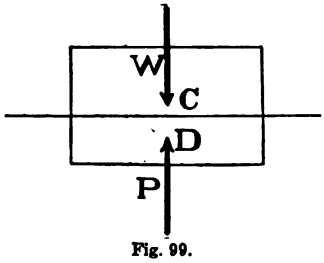


Fig. 99.

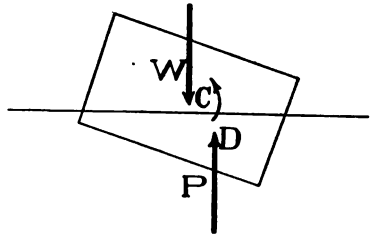


Fig. 100.

The condition of equilibrium is generally described by the use of a particular point, called the *metacenter*. If the body is given an infinitesimal inclination (Fig. 101), the point M at which the axis is intersected by the resultant of the upward pressures is the metacenter. The body is in stable equilibrium when the metacenter lies above its center of gravity. In this case, whenever the body is inclined, its weight and the resultant of the upward pressures combine to form a restoring couple.

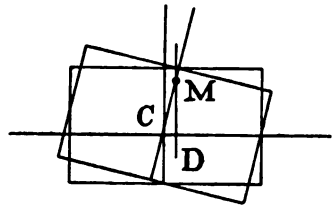


Fig. 101.

The height of the metacenter above the center of gravity is called the *metacentric height*. The degree of stability is determined by the magnitude of the metacentric height.

126. Relative Density or Specific Gravity. — If we weigh equal volumes of different substances we find that in general their weights are different. When the volume which is chosen is some standard volume, these different weights are characteristic of the different substances.

The mass of a substance which is contained in unit volume is called the *density* of that substance. If we can prepare portions of different substances of unit volume or of any known volumes, we can determine their masses by weighing them, and thus determine their densities.

It is a very difficult task to obtain bodies whose volumes are accurately known, and we accordingly are more apt to compare substances in another way. Instead of choosing a standard volume with which to determine densities directly, we choose a standard substance, and determine the ratio of the density of any other substance to the density of the standard substance. This ratio we call the *relative density* of the substance. The standard substance generally chosen is water.

The relative density is also often called the *specific gravity* of the substance, and when so called it is generally defined as the ratio of the weight of the substance occupying a certain volume to the weight of water occupying an equal volume.

If we represent the mass of a body of any particular substance by M_1, and its volume by V_1, then its density D_1 is given by $D_1 = \frac{M_1}{V_1}$.

The density of water is given by a similar expression, $D_0 = \frac{M_0}{V_0}$.

According to the definition of relative density, the relative density S_1 of the substance chosen is

$$S_1 = \frac{D_1}{D_0} = \frac{M_1 V_0}{M_0 V_1}.$$

If $V_1 = V_0$, we have $S_1 = \frac{D_1}{D_0} = \frac{M_1}{M_0}$, the latter ratio defining the specific gravity of the substance.

When the metric system of weights and measures was introduced, the attempt was made by Borda to construct the standard mass, or kilogram, equal to the mass of a cubic decimeter of water under standard conditions. In this attempt he succeeded very nearly, so that unless extraordinary accuracy is desired, the mass of a quantity of water in kilograms is numerically equal to the volume of the water in cubic decimeters. The gram is the mass of water in a cubic centimeter. In the system of units in which the gram is the unit of mass, the centimeter the unit of length, and the cubic centimeter the unit of volume, the density of water is unity, and therefore the density of any other substance is expressed by the same number as its relative density. In other systems of units this is generally not the case. The relative density, which is a simple ratio between two densities, is the same for all systems of units; the absolute density is different in the different systems.

127. Determinations of Relative Density. — A vessel may be arranged full of water, so that the water which overflows when a body is immersed in it can be caught and weighed. The volume of the water which overflows is equal to the volume of the body, and the ratio of the weight of the body to the weight of the water is therefore the relative density.

The density of solids which are insoluble in water is commonly obtained by the use of the *hydrostatic balance*. This is a balance so constructed that the body to be tested can hang below one of the scale pans. The body is first weighed in air. It is then immersed in water and its apparent weight determined. By Archimedes' principle, the loss in weight which it apparently undergoes is the weight of an equal volume of water. The ratio of the body's weight to this loss in weight is therefore the relative density of the substance of the body.

The relative density of a liquid is often determined by an instrument called the *hydrometer*. Hydrometers are of two kinds, those of constant weight and those of constant volume.

The *hydrometer of constant weight* is generally a glass bulb weighted at the bottom, and furnished above with a long cylindrical stem. When it is placed in a liquid it stands upright in it, and a proper adjustment of its weight will cause it to sink so that

part of its stem is immersed. It is standardized by placing it in water and marking the point on the stem at which the stem protrudes from the surface of the water. It is then placed in another liquid of known density, and the place on the stem again marked at which it protrudes from the surface of the liquid. The distance between these two points is marked off into any arbitrary number of equal divisions, and this graduation is extended above and below the standard points. When the instrument, thus prepared, is placed in any liquid, it sinks to a certain point, which may be read off by means of the graduation. Its weight remaining always the same, the weight of the liquid which it displaces is always the same, and since the point to which it sinks determines the volume of the liquid displaced, we have the means of determining the volumes of equal weights of different liquids, and so of determining their relative densities.

The *hydrometer of constant volume* is a glass or metal bulb, weighted at the bottom, and carrying at the top a cylindrical stem supporting a pan in which weights may be placed. The instrument is floated in a liquid and weights are placed in the pan until the instrument sinks to a marked point on the stem. The weight of the instrument and the weights in the pan are together equal to the weight of the liquid displaced. By using this instrument in different liquids, one of which is the standard liquid, we may determine the weights of equal volumes of the liquids, and therefore their relative densities.

The relative density of liquids is often determined by the use of the *specific gravity bottle*, or *pyknometer*. This is a small bottle or flask or other vessel of constant volume. The weight of water which will fill it is determined once for all. To determine the relative density of another liquid, it is filled with that liquid and the weight of the liquid determined.

128. The Barometer. — The attention of Galileo was once called to the fact that the water of a certain deep well could not be drawn out of it by a pump. The water rose in the pipe about 34 feet and could not be raised further by continued pumping. At that time the rise of water in a pump was considered an exhibition of a general principle, embodied in the statement "Nature abhors a vacuum." Galileo recognized that the fact which had been called to his attention proved at least that this principle was limited in its application, but he was not able to explain it.

Torricelli, one of his friends, starting with the knowledge which he had obtained from Galileo that air has weight, considered it to be a simple case of equilibrium between the pressure of the water column in the pipe and of the atmosphere. To verify this conclusion, he filled a long glass tube, closed at one end, with mercury, and stopping the other end with his finger, inverted the tube and inserted the lower end of it in a vessel of mercury. On removing his finger, the column of mercury in the tube fell until its highest point was about 30 inches above the surface of the mercury in the vessel and settled to rest in that position. On the supposition that the mercury column is sustained by the pressure of the atmosphere on the free surface of the mercury in the vessel, the height of the mercury column ought to be proportioned to the height of the water column, in the experiment with the pump, inversely as the respective densities of mercury and water. For the pressure of the atmosphere is the same in both cases, and therefore the pressures of the two columns of mercury and water should be the same, if the hypothesis is correct.

The general expression for the pressure of a column of liquid may be found by considering a column standing on an area a, and of height h. The mass of this column is equal to its volume multiplied by the density D of the liquid. We have therefore the weight of the column $W = Dgah$, and hence the pressure at the area a is

$$p = \frac{W}{a} = Dgh. \tag{64}$$

Two liquid columns will exert the same pressure when this product Dgh is the same for both of them. Their heights are therefore inversely as their densities.

To apply this to the case under consideration we need, in addition to the data already given, the fact that the density of mercury is 13.6 times as great as the density of water. Accordingly, the height of the water column sustained by the pressure of the atmosphere should be 13.6 times as great as the height of the corresponding mercury column. This conclusion from Torricelli's hypothesis was verified by his observations, and he accordingly concluded that the water rose in the pump and that the mercury was sustained in the tube by a common cause, the pressure of the atmosphere.

The arrangement constructed by Torricelli may be set up permanently as a means of measuring the pressure of the atmosphere.

It is then called a *barometer* (Fig. 102). The pressure of the atmosphere is found to vary from time to time, but it never differs very much at any place from a mean or standard pressure for that place.

Fig. 102.

Pascal contributed to the verification of Torricelli's hypothesis by showing that the height of the mercury column in the barometer diminishes when the barometer is carried up a mountain or to the top of a high tower. The reason for this is plain when we notice that the barometer at the higher station is relieved from the pressure of a column of air, whose height is the difference of level between the lower and upper stations.

The pressure indicated by the barometer under standard conditions is often used as a unit of pressure. This unit of pressure is called the *pressure of one atmosphere*, or sometimes the atmo. It is the pressure indicated by the barometer when placed at the sea level and when the height of the column is 760 millimeters. Taking the density of mercury as 13.6, the standard height of the mercury column as 76 centimeters, and g as 980, we get for the pressure of one atmosphere in c.g.s. units the value 1.013×10^6 dynes per square centimeter.

Hydrodynamics

129. Steady Motion. — The determination of the motion of a liquid as it flows from one place to another is in general a task of extreme difficulty, and success cannot be reached without the use of advanced mathematics. There is one general type of fluid motion, however, called steady motion, which can be handled more simply, by the aid of the principle of the conservation of mechanical energy.

In *steady motion* the liquid remains continuous, and the portions of it which pass successively through any particular point in space all have the same velocity and move in the same direction. If any one portion is followed from one end of its path to the other, it will traverse a line, called a *stream line*, and the portions which start from the same point, and follow it, will all traverse the same line. If two surfaces are drawn in the liquid perpendicular to the stream lines which pass through them, the portions of liquid which are contiguous to each other as they cross the surface through

which they enter, will flow along contiguous stream lines, and will emerge through the other surface together. Such a motion is approximately represented by the motion of water flowing under pressure through a series of pipes.

130. Condition of Continuity. — Let us consider a liquid flowing with steady motion. On one of two surfaces which are drawn perpendicular to the stream lines, describe a closed curve. The stream lines which pass through the points of that curve will form a tubular surface, and will pass through the other surface in contiguous points, which will also trace a closed curve. We may call the mass of moving liquid within this tube a *stream* (Fig. 103).

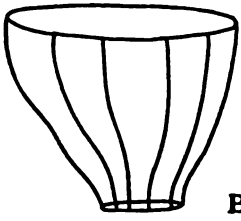


Fig. 103.

Because of the condition that the motion is steady, as much liquid will enter in a unit of time through the surface A as flows out through the surface B. If u represents the average velocity of the portions of the liquid which pass through A, and v that of those portions which pass through B, the volume of liquid that will enter through A in the small time t will be Aut, while that which flows out through B will be Bvt. Since these volumes are equal if the liquid is practically incompressible, we have $Au = Bv$; and this relation holds true for any two surfaces which may be drawn perpendicular to the stream lines. We may call it the *condition of continuity*.

131. Torricelli's Theorem. — It was observed by Torricelli that when a stream of water issues from a small orifice in the side of a vessel full of water, the velocity of the stream is the same as that which a body would acquire by falling from the surface of the water to the orifice. This relation is called *Torricelli's theorem*. It may be tested experimentally by fitting to the orifice a short tube so bent that the stream issuing from it is directed upwards, in which case it will be found that the stream rises to the same level as that of the water in the vessel. It may also be tested by allowing the stream to issue horizontally and determining the dimensions of the parabola which it describes.

When we attempt another verification of Torricelli's statement, by a determination of the quantity of water which issues in a given time, we obtain results inconsistent with it. The amount of water obtained is always less than that which should issue from an orifice of the given size, if the velocity is that stated in Torri-

celli's theorem. Newton observed that, owing to the way in which the parts of the water near the orifice rush together as they issue from it, the diameter of the stream outside the orifice is not so great as that of the orifice. The narrow portion of the stream is called the *vena contracta*. When the quantity of water which should issue is calculated as if it were issuing from an orifice whose diameter is that of the *vena contracta*, the experimental results obtained are in better agreement with the calculation.

We may deduce Torricelli's theorem by the aid of the principle of the conservation of mechanical energy. Let us consider a large vessel (Fig. 104) furnished with a small orifice in the bottom. It is filled with liquid up to a certain level, and as the liquid runs out from the orifice the level is maintained by admitting an equivalent amount at the top. After a while the flow will become steady. The energy of the mass of liquid contained between the upper surface and a surface drawn across the orifice will thereafter remain constant, and consequently, by the principle of conservation, the energy of the liquid which enters at the top will equal that which emerges at the orifice. Referring the potential energy of the liquid to the level of the orifice, calling the mass which enters or leaves the vessel in unit time m, the velocity in the top layer u, the velocity of efflux v, and the height of the upper level above the orifice h,

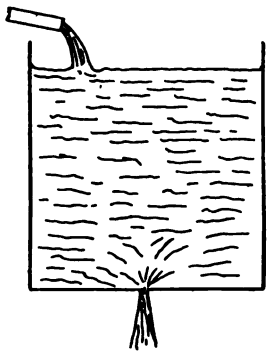


Fig. 104.

we have the potential energy of the liquid at the top given by mgh, its kinetic energy by $\frac{1}{2}mu^2$, and the kinetic energy of the outgoing mass by $\frac{1}{2}mv^2$. Setting the entering energy equal to the outgoing energy we have

$$mgh = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mu^2.$$

If we represent by a and o the respective areas of the upper surface and of the orifice, the condition of continuity (§ 130) gives us $au = ov$; and using this relation to eliminate u from the equation of energy, we obtain

$$2gh = v^2\left(1 - \frac{o^2}{a^2}\right).$$

This equation for the velocity reduces to that stated in Torricelli's theorem, if the area of the orifice is negligibly small in comparison with that of the upper surface; for then

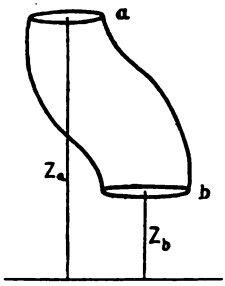
$$v^2 = 2gh. \qquad (65)$$

It is worth noticing that the velocity of efflux in this case is independent of the density of the liquid.

132. Bernoulli's Theorem. — The theorem which has just been deduced is a particular case of a more general theorem descriptive of the steady motion of a liquid, due to Daniel Bernoulli. It was published by him in his *Hydrodynamica* (1738), the first important work treating of the motions of liquids.

To prove Bernoulli's theorem, we must first consider the way in which the pressure in a liquid does work on the moving parts of the liquid. Let us think of a surface drawn perpendicular to the stream lines, after steady motion is established, and consider so small a portion of it that the pressure p at every part of it is the same. If a represents its area, the force exerted upon it in the direction in which the stream is flowing is pa; and if we imagine the surface to move forward with the stream for the small distance s in the small time t, the work done by the pressure is pas, and the rate at which it is done is pas/t. But s/t is equal to the velocity v, and therefore the rate of work done at the surface where the velocity is v and the pressure is p, on the liquid which passes through the area a, is pav.

Now to prove Bernoulli's theorem, let us consider a stream of liquid, limited as before (Fig. 105) by two surfaces perpendicular to the stream lines; and let us suppose the cross sections of the stream so small that the pressure and velocity are the same at all points of each of these surfaces. Represent the pressure at the surface of area a by p_a, that at the surface of area b by p_b; and the respective velocities at those surfaces by v_a and v_b. Let the plane of reference for the potential energy of any part of the liquid due to its weight be so placed that the points of the area a are above it at the distance z_a; those of area b above it at the distance z_b. Now $v_a at = v_b bt$ equals the volume V which enters or leaves the region in the time t. The excess of work done on the entering mass over that done on the outgoing mass may therefore be written $(p_a - p_b)V$. Since the motion is steady and the total energy in the stream is constant, this excess of work done on the entering mass must be equal to the excess of energy, potential and kinetic, which leaves the stream with the outgoing mass above that which enters with the entering mass.



[Fig. 105.

The mass which enters the region through the first surface and leaves it through the second surface may be represented by ρV, if we use ρ to represent the density of the liquid. The potential energy which enters with the entering mass is $\rho V g z_a$, that which leaves with the outgoing mass is $\rho V g z_b$. The kinetic energy which enters with the entering mass is $\frac{1}{2}\rho V v_a^2$, that which leaves with the outgoing mass is $\frac{1}{2}\rho V v_b^2$.

Since the excess of work done is equal to the excess of energy, we have

$$(p_a - p_b)V = \rho V g(z_b - z_a) + \tfrac{1}{2}\rho V(v_b^2 - v_a^2),$$

from which we obtain

$$p_a + \tfrac{1}{2}\rho v_a^2 + \rho g z_a = p_b + \tfrac{1}{2}\rho v_b^2 + \rho g z_b. \qquad (66)$$

This equation is one form of Bernoulli's theorem.

To deduce Torricelli's theorem from it, we notice that the pressures p_a amd p_b on the free upper surface and at the orifice respectively are equal, being both equal to the atmospheric pressure. The distance $z_a - z_b = h$, the height of the free surface above the orifice. The velocity at the upper surface is negligible. Hence Bernoulli's formula in this case reduces to

$$2g(z_a - z_b) = 2gh = v_b^2,$$

which is the formula found directly in the discussion of Torricelli's theorem.

133. Pressure in a Stream. — In a stream of liquid in which the conditions are such that we can ignore the effect of the weight of the liquid, Bernoulli's theorem becomes

$$p_a - p_b = \tfrac{1}{2}\rho(v_b^2 - v_a^2).$$

This means that at the surface a the pressure will be greater or less than that at the surface b, according as the velocity at b is greater or less than that at a. By the help of the condition of continuity, we may put the equation in the form

$$p_a - p_b = \frac{1}{2}\rho v_a^2\left(\frac{a^2}{b^2} - 1\right); \tag{67}$$

and considering the pressure and velocity at a as constant we can trace the changes in pressure as we pass along the stream from one cross section to another. In general, when a is greater than b, p_a is greater than p_b, so that the pressure is less where the cross section is less. A stream of water flowing through a pipe, or a series of pipes, with varying cross section will exert less pressure against the walls of the pipe in those parts where the column is narrower and the stream is flowing swifter. The pressure may fall off so much as to be less than the atmospheric pressure. In this case, if a small hole is made in the side of the pipe, water will not flow out, but air will enter and will be carried on with the water.

134. Vortex Motion. — If we conceive of a line drawn in a fluid so as to determine a particular set of fluid particles or, more properly, elements, and if these elements are rotating about this line as an axis, the assemblage of such elements constitutes a vortex line. Strictly speaking, vortex lines in the mathematical sense cannot exist in a real fluid, whose particles, though small, are of finite dimensions, but vortices possessing approximately the properties demonstrated by analysis for the true vortex can be set up and examined by experiment.

It was proved by Helmholtz that the strength of the vortex, that is, the product of the angular velocity of an element of the vortex around the axis and the cross section of the vortex, is constant along the vortex line. As a consequence of this, the vortex cannot end within the fluid. The vortex line is either a closed curve, or ends on the surfaces of the fluid. Further, the portions of the fluid which are in the vortex originally always remain in it, so that the vortex proceeds through the fluid, not as a form of motion, like a wave, but like a body, being a special portion of the fluid, distinguished from the rest of it by a peculiar motion

and moving as a whole. An example of the closed vortex is seen in a smoke ring. The open vortex, with both ends at the same surface, is often seen when an oar blade is drawn through water. The open vortex with its ends at different surfaces is probably exemplified by the cyclone or whirlwind.

In a frictionless fluid the vortex cannot be set up or destroyed by any finite forces. Two vortex lines cannot cut each other, and hence if the line of a vortex is knotted or implicated on itself, the degree of implication will always persist in it. Finally the existence of a vortex generally involves its motion through the fluid and also the existence of currents in the surrounding fluid. For example, in the case of a circular or ring vortex, like a smoke ring, the vortex moves forward perpendicularly to the plane of the ring, and the fluid in the middle of the ring moves forward more rapidly even than the ring does. The motions of neighboring vortices are affected by these currents, so that the vortices are accelerated as if forces were acting between them.

In actual fluids, friction soon alters the vortices which can be set up and finally destroys them. It is also true, on the other hand, that if it were not for friction it would not be possible to set up even the imperfect vortices which can be made.

The interest in the properties of vortices was until recently peculiarly great, because it was thought that, as suggested by Lord Kelvin, the properties of the atoms of matter could be explained by supposing them to be vortices in some universal and infinitely extended frictionless fluid. The atoms were supposed to be, like the vortices, indestructible individuals, which are in continual motion, and which act on each other by forces. It was not easy, on the vortex-atom theory, to explain inertia, but practically all the other properties of matter were fairly well accounted for. The theory has been abandoned since it has been shown, by the study of radium and other similar substances, that the atoms of matter are not the indestructible and indivisible units which they were believed to be, and which they would have to be if they were vortices in a frictionless fluid.

135. Waves on the Surface of a Liquid. — When a liquid surface is disturbed, the disturbance passes outward in every direction as a wave or a series of waves. Each of these waves contains a portion of the liquid which is elevated above the general level, and another portion which is depressed below it. Our common observation of the movements of floating bodies when waves pass

under them, shows that the liquid is not carried along with the waves. The wave is therefore simply a mode of motion which is impressed successively upon the different parts of the surface. We may study the motion of the liquid caused by the waves, by immersing in it small fragments of some solid whose relative density is the same as that of the liquid. In the case of water, small fragments of amber may be used.

Experiments on this plan were carried out by E. H. Weber and W. Weber (1825), in a long shallow trough with plane vertical glass sides. They found as a general rule, for a shallow body of liquid, that if a succession of similar waves passes over the surface, the particles in the surface describe circles, whose diameters are equal to the vertical distance between the highest point and the lowest point of the wave. At the highest point of the wave they move forward in the direction in which the wave is traveling. At the lowest point they move backward with the same velocity. The fragments which are below the surface describe ellipses, which become smaller and smaller, and more and more eccentric as the depth increases, until at last they cannot be distinguished from minute horizontal lines.

The disturbance set up by waves on the surface does not penetrate far beneath it, and when the depth of the liquid is very great, most of it remains quiet, even when the waves which pass over it are large. In this case the theory of the motion shows that

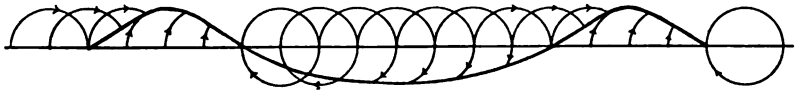


Fig. 106.

for low waves the paths of all the particles are circles (Fig. 106), which decrease in diameter as the depth increases. As the diagram shows, the troughs of these waves are longer than the crests. When the waves are high, the form of the wave can no longer be described by such simple circular motions. The trough of the wave becomes

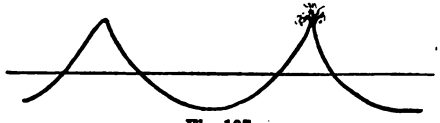


Fig. 107.

relatively longer still (Fig. 107), and the crest often terminates abruptly at the top in a sharp angle, or the liquid is so thrown up

at that point that the wave breaks. When such waves advance from a deeper to a shallower part of the liquid, the backward part of the motion, which is in the lower part of the curves described by the parts of the liquid, is retarded, and the forward motion is

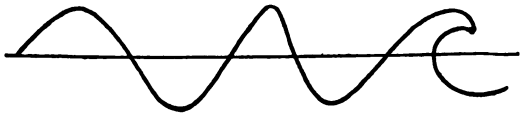


Fig. 108.

in excess (Fig. 108). The top of the wave thus moves forward faster than the bottom, and if the wave advances in such a direction that the depth continually decreases, it will at last curl over and break.

BOOK II

GENERAL PROPERTIES OF MATTER

CHAPTER V

GRAVITATION OR MASS ATTRACTION

136. Kepler's Laws of Planetary Motion. — By an extended study of the apparent motions of the planet Mars, and then of those of the other planets, Kepler showed (1609–1619) that they could be represented by supposing the motions of the planet to conform to the following laws:

1. The path of a planet is an ellipse, with one of its foci situated in the sun.

2. The radius vector drawn from the sun to a planet sweeps out equal areas in equal times.

3. The squares of the periodic times of the planets are proportional to the cubes of the semi-major axes of their elliptic orbits.

By the *periodic time* of a planet is meant the time required by it to describe its orbit.

137. The Problem of the Planetary Motions. — The laws of Kepler merely describe motions which are consistent with the apparent motions of the planets. On the assumption that they describe the real motions of the planets, it became a question of the utmost interest to determine the forces which caused the motions. This question was first answered by Newton in the *Principia* (1687). Newton's investigation is an excellent illustration, perhaps the best illustration which we have, of the process of thought employed in physical reasoning, as described in § 2.

Newton suspected that each planet is acted on by a force which is proportional to its mass, is directed toward the sun, and is inversely proportional to the square of the distance between the planet and the sun. He therefore applied the principles of mechanics to determine what would be the motion of a body acted on by a force directed toward a fixed point according to this law.

138. Attraction of Spheres. — Newton proved that, if such a law of attraction holds true, not only for the planets and the sun, but for all particles of matter, a sphere of matter will attract a small body outside of it just as if its whole mass were concentrated at its center (Examples, XI, 2). He was thus entitled to deal with the planets and the sun in his calculations as if they were material particles or small bodies.

139. Law of Equal Areas. — Newton then showed that if the force which acts on a body is directed toward a fixed point, whatever may be the way in which the force varies with the distance of the body from that point, the radius vector drawn from the point to the body will sweep out equal areas in equal times. That is, a body acted on by a force which is always directed toward a fixed center conforms to Kepler's second law. Hence the conclusion may be drawn that the force which acts on a planet is directed toward the sun.

Let a body be moving through the point A with the velocity v_0 (Fig. 109). In the small time t it will pass over a distance s from A to B, given by $s_0 = v_0 t$.

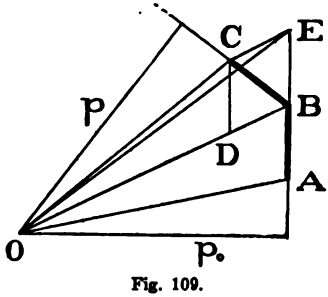


Fig. 109.

At B let it receive an impulse directed toward the fixed point O. Its motion in the next following equal time t will be along a line BC, which is the resultant of the line BE, equal to AB, which it would have traversed if the impulse had not been applied, and the line BD, directed toward O, which it would have traversed if its motion had been solely that given it by the impulse. The distance BC traversed by it will be $s = vt$. The areas swept out by the radius vector in the equal times are the areas of the triangles OAB and OBC. But these triangles are equal; for $OAB = OBE$, being on equal bases AB, BE, and having a common apex at O; and $OBE = OBC$, being on the same base OB, and between the same parallels OB and CE.

If p_0 and p represent perpendiculars from O on the lines of motion AB and BC, the areas of the triangles OAB, OBC are given by the products of their bases and these perpendiculars, and hence $p_0 v_0 t = pvt$, or $p_0 v_0 = pv$.

If another impulse toward O is applied to the body at C, a similar equality may be proved; and if the motion is not effected by impulses, but by a continuous force directed toward O, the same proof applies if the distances AB, BC, etc., are taken as the infinitesimal elements of the curved path of the body. The direction of motion is at any point the direction of the tangent to the path, and p and p_0 are perpendiculars from the fixed center to these tangents. For any two points in the path we have

$$p_0 v_0 = pv. \tag{68}$$

140. Motion in a Conic Section. — Newton showed further that when the force which acts on the body varies inversely with the square of the distance between the body and the fixed center, and is directed toward that center, the path of the body will be one of the conic sections. Whether the path of the body is an ellipse, a parabola, or a hyperbola, depends on the velocity of the body as it passes the end of the major axis. When this velocity does not exceed a certain limit, the path of the body will be an ellipse. The motion of the body, therefore, which is caused by the hypothetical force supposed to act on it, conforms to Kepler's first law. Hence the conclusion may be drawn that the force which acts on a planet toward the sun varies inversely with the square of the distance between the planet and the sun.

Consider a body of mass μ moving in a path under the influence of such an attraction toward the fixed center O (Fig. 110). The force with which it is attracted toward O may be represented by $F = m/r^2$, if r is the distance from the body to the fixed point, and m is a factor of proportion. The component of that force along the perpendicular to the path is $F\cos\alpha$, and $\cos\alpha = p/r$. This perpendicular component is the centripetal force which determines the curvature of the path. If ρ represents the radius of curvature at A, the centripetal force is equal to $\mu v^2/\rho$. Setting these forces equal we have

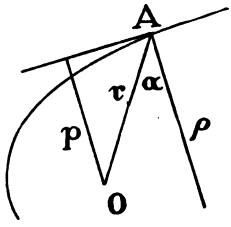


Fig. 110.

$$\frac{\mu v^2}{\rho} = \frac{mp}{r^3}.$$

Now from the theorem of the last section, if p_0 and v_0 represent values of p and v which occur together at some particular instant, we have $pv = p_0v_0$, and substituting the value of v, we obtain the equation $\frac{(p_0v_0)^2}{\rho} = \frac{mp^2}{\mu r^3}$, and hence

$$\rho = \frac{(p_0v_0)^2\mu}{m}\frac{r^3}{p^3}, \tag{69}$$

as a formula for the radius of curvature of the path at any point in terms of the radius vector from the fixed point and the perpendicular on the tangent. Now this formula is a formula for the radius of curvature of a conic section; and hence the path given the body by a force acting according to the assumed law is one of the conic sections.

For each of the conic sections there is one point at which the tangent is perpendicular to the radius vector, so that $r = p$. If we take this point for that at which p_0 and v_0 are determined, and call the radius of curvature at that point ρ_0, we have $\rho_0 = \frac{(p_0v_0)^2\mu}{m}$. Now the radius of curvature of the ellipse at the end of the major axis is b^2/a, if a and b represent the semi-major and semi-minor axes respectively; that of the hyperbola is also b^2/a; and that of the parabola

is $2d$, if d represents the distance from the focus to the vertex. We therefore have for the ellipse and hyperbola, $\frac{(p_0v_0)^2\mu}{m} = \frac{b^2}{a}$; and for the parabola $\frac{(p_0v_0)^2\mu}{m} = 2d$.

In the equation for the ellipse and hyperbola we may eliminate the b and the a by the use of the eccentricity e, using the formulas, $p_0 = \pm a(1-e)$, $e = \sqrt{\frac{a^2 \mp b^2}{a^2}}$; and obtain

$$\mu v_0^2 = \frac{m}{p_0}(1+e). \tag{70}$$

If we fix the value of the length p_0, the distance from the center of force to the vertex of the conic, the value of e will depend upon the value of v_0. If $\frac{\mu v_0^2}{2} = \frac{m}{p_0}$, then $e = 1$, and the conic is a parabola. If $\frac{\mu v_0^2}{2} \lessgtr \frac{m}{p_0}$, then $e \lessgtr 1$, and the curve is either an ellipse or a hyperbola. The term $\mu v_0^2/2$ is the kinetic energy of the body at the vertex of the conic, and it will be shown later (Examples, XI, 7) that m/p_0 measures the work that would be done by the force on the body, if the body were to travel under it from an infinite distance to the vertex; in other words, it measures the potential energy lost by the body in moving from infinity to the vertex. Whether the path of the body shall be an ellipse, a parabola, or a hyperbola therefore depends on whether its kinetic energy at the vertex is less than, equal to, or greater than the potential energy which it would gain by removing to infinity.

141. The Periodic Times. — Newton showed further that when forces directed toward a common center and varying inversely with the square of the distance act on different bodies so as to make them describe elliptic orbits, the squares of the times in which they describe their orbits are proportional to the cubes of the semi-major axes of the orbits, if the force which acts on each body is proportional to the mass of the body. That is, if the forces which act upon the bodies are proportional to their masses, the motions of the bodies will conform to Kepler's third law. Hence the conclusion may be drawn that the forces which act upon the planets are proportional to their masses.

In the last section we found for the ellipse $\frac{(p_0v_0)^2\mu}{m} = \frac{b^2}{a}$. In § 139 we found that the rate at which the radius vector sweeps out area, or the areal velocity, is given by p_0v_0. If we represent by T the periodic time of the motion of a body in an ellipse, this areal velocity is also given by $\pi ab/T$, since πab is the area of the ellipse, so that we may set $p_0v_0 = \frac{\pi ab}{T}$. Using this value for p_0v_0 in the equation above, we obtain

$$m = \frac{\pi^2 a^3 \mu}{T^2}. \tag{71}$$

If therefore, for different bodies moving in elliptic orbits around the same center, the ratio a^3/T^2 is constant, the factor of proportion m for each body is proportional to its mass. But the distances being equal, the factors of proportion are proportional to the forces, so that the forces are proportional to the masses.

142. Force of Gravitation. — The results of the preceding paragraphs, which may be summed up in the statement that the planets move as if they were attracted toward the sun by forces which are proportional to their masses and which vary inversely with the squares of their distances from the sun, are all that can be reached by the study of the motion of the planets as expressed in Kepler's laws. More might have been learned from a study of the perturbations caused in the orbit of one planet by the action of the others around it. Newton, however, did not proceed along this line in his argument, but followed another course. He showed that the motion of the moon conforms to Kepler's first and second laws when the earth is taken as the center of force, and therefore concluded that the moon is acted on by a force directed toward the earth and varying inversely with the square of the distance of the moon from the earth. He then showed, from a calculation of the dimensions of the moon's orbit, that the acceleration of the moon toward the earth is equal to the acceleration of a body near the earth's surface diminished in the inverse ratio of the squares of the distances of the moon and of the body from the center of the earth. From this the conclusion was drawn that the force which acts on the moon and determines its motion in its orbit is of the same nature as the force which causes heavy bodies to fall toward the earth; and further, that the forces which act between the different bodies of the planetary system are of the same nature. This force is therefore called the *force of gravitation.*

This step in Newton's series of proofs is essentially the same thing as proving that the moon's motion obeys Kepler's third law. For if a body could be projected horizontally from some point near the earth and given sufficient velocity, it would move round the earth in a circular orbit, and its centrifugal force would be equal to its weight. It would become another satellite of the earth. Its motion and that of the moon will be found to conform to Kepler's third law.

143. Proportionality of Mass and Weight. — Newton constructed two pendulums of equal length whose bobs were two similar spherical boxes in which different bodies could be placed. He found that whatever were the substances placed in the boxes, the times of oscillation of the pendulums were the same, and hence concluded that the force of gravity is proportional exactly to the mass of the body acted on, and is independent of its other characteristics.

Our determination of the law of the pendulum was made on the assumption that the mass and weight of the bob were propor-

tional. On that assumption the mass falls out of the equation of motion, and the time of oscillation depends only on the length of the pendulum. This being shown to be in accord with experiment, the assumption of a proportionality between mass and weight is verified.

144. Mutual Attractions of Masses. — Observations of the motions of Jupiter's satellites, as well as of the satellites of the other planets, show that they conform to Kepler's third law when the planet is taken as the center of force, and hence that the forces which act on them toward the planet are proportional to their masses. On the assumption that the force with which the planet acts on the sun follows the same law, the attraction between the sun and the planet is proportional to the mass of the sun as well as to the mass of the planet.

145. Law of Gravitation. — The arguments which have now been adduced led Newton to the conclusion that a force acts between any two portions of matter in the universe, which tends to draw them together, and that the magnitude of this force is proportional to each of the masses and inversely proportional to the square of the distance between them. This law is called the *law of gravitation* or, from the fact that the force depends directly on the masses of the interacting bodies, the *law of mass attraction*. If we represent by m the mass of one of the bodies, by m' the mass of the other, by r the distance between them, and by k a constant or factor of proportion, we may express the force of gravitation F between these bodies by the formula

$$F = k\frac{mm'}{r^2}. \tag{72}$$

The factor of proportion k is the same in all cases.

146. Gravitation Constant. — The factor of proportion which has been introduced in the formula just given is called the gravitation constant. It is manifestly equal to the force which two unit masses will exert on each other when they are unit distance apart. It may be determined by a direct determination of the gravitational force with which two bodies, whose masses are known, attract each other.

An experiment designed for this purpose by Michell was carried out by Cavendish in 1798. It is always known and referred to as *Cavendish's experiment*. A horizontal rod (Fig. 111), carrying on its ends two equal balls of lead, was suspended at its middle point from a long cylindrical wire. When not acted on by any other

forces than its own weight, such a system will assume a definite position, in which there is no twist in the wire. If the wire is twisted by turning the rod through any angle in the horizontal plane, the twist in the wire will introduce an elastic reaction proportional to the angular displacement of the rod, and by preliminary experiments the magnitude of the couple which will cause a given twist in the wire may be determined. After such a determination had been made, two large balls of lead were so placed on either side

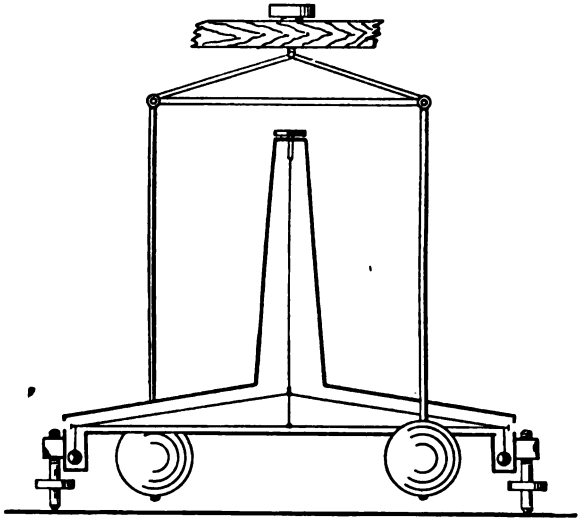


Fig. 111.

of the rod as to attract the small balls in opposite directions, and so to introduce a couple which twisted the wire. The magnitude of this couple was measured for different distances between the large and small balls, and thus the forces which they exerted on each other were determined. Cavendish found that the force between the balls varied inversely as the square of the distance between their centers, and, by using balls of different masses, that the force between them was proportional to the masses. The values of the forces thus determined, introduced, with the known masses of the balls and the distance between their centers, into the formula expressing the law of gravitation, led to the determination of the gravitation constant.

The Cavendish experiment has been several times repeated, with every precaution to insure accuracy. The results obtained by different observers are fairly consistent with one another. The value

of the gravitation constant obtained from them, when the unit of mass is the gram and the unit of distance the centimeter, is about sixty-five billionths of a dyne.

That is, if in the equation $F = k\frac{mm'}{r^2}$, we set $k = 0.000000065$, and express m and m' in grams and r in centimeters, the value calculated for F will give the gravitational force, expressed in dynes, between the two masses m and m' when separated by the distance r.

A method employed by von Jolly led to the same result. In this method a sensitive balance was mounted high in a tower and additional scale pans were suspended from it by long wires. This balance was first used to show that the weight of a body varies with its distance from the earth's surface. A weight placed in one of the upper scale pans was balanced by another weight in the other upper scale pan. This second weight was then placed in the lower scale pan and found to be heavier than before. Its increase in weight was determined, and measured the increase in the attraction of the earth upon it, due to its being brought nearer the earth's center. To determine the gravitation constant, a large block of lead was placed under the weight in the lower scale pan, which, by its attraction, caused a further increase in that weight. This was carefully measured, and from it, taken together with the known values of the attracting masses and their dimensions, the gravitation constant was calculated.

By using an improved form of von Jolly's experiment, Richarz and Krigar-Menzel have found for the constant of gravitation, $k = (6.682 \pm 0.011) \cdot 10^{-8}$.

147. Acceleration at the Earth's Surface Due to Gravity. — In studying the motions of bodies caused by the earth's attraction we may consider the earth as fixed, the movement of the earth toward the common center of mass, which occurs when the body moves toward the earth, being so small that it may be neglected. This being so, it follows from the law of gravitation that all bodies near the earth's surface will have the same acceleration g. As has already been seen, this conclusion was reached by Galileo from his observations. It is of great practical importance to know its exact value, because it furnishes us the most precise measure of force in absolute units. We can determine the mass of a body very exactly, and use its weight as a force. Then the force exerted by the body, expressed in absolute units, is its mass multiplied by this acceleration.

Bodies fall so fast that it is extremely difficult to make exact observations on them. A method has been devised for this purpose, in which the time of fall is recorded electrically, which yields fairly accurate values of g. Observations on bodies rolling down an inclined plane, or with the Atwood's machine, will yield approximate values of this constant. The measurements of g which are sufficiently exact to be of any scientific value are made universally by means of the pendulum, first used for this purpose by Huygens.

In observations with the pendulum, two forms are used, of which the theory has already been given in §§ 105, 106. In one of these, known as Borda's pendulum, the bob is a heavy sphere and the suspension a long cylindrical wire. The wire is hung from a knife-edge, which rests on agate planes. The effect of the knife-edge on the period of the pendulum is avoided by a preliminary adjustment. The efficient parts of the pendulum are therefore regular bodies and the moment of inertia and the static moment of the pendulum can be calculated. The period of oscillation is observed, and, by the aid of the formula given in § 105, the value of g is determined.

In the other form, known as Kater's pendulum, the pendulum is a stiff rod or bar carrying a heavy weight at one end and furnished with two knife-edges which confront each other. An additional movable weight is also carried on the bar. By a suitable adjustment of this weight the periods of the pendulum swinging about the two knife-edges may be made the same. The pendulum thus adjusted is a reversible pendulum, and, as was seen in the discussion of § 106, the distance between the knife-edges is the length of the simple pendulum which will swing in the same period. Having measured this distance once for all, and having determined the period, the value of g may be calculated by the aid of the formula for the simple pendulum.

The value of g is different at different places, ranging between 978 at the equator and 983 at the pole, when the units of length and time are the centimeter and second. For ordinary calculations in our latitude we may use the value 980.

148. Density of the Earth. — Newton did not attempt any experimental demonstration of his hypothesis that an attraction exists between any two bodies. The experiment of Cavendish by which this hypothesis was confirmed was made long after his time. The attempts which were made to confirm it, before Cavendish's experiment was executed, involved a comparison of the attraction of the earth for a body with the attraction exerted on that body

by a portion of the earth, and the numerical results obtained were worked over so that the result stated was a value of the mean density of the earth.

The first of these experiments was made by the astronomer Maskelyne in 1774. The theory of this experiment is as follows: If two stations on the same meridian are chosen and the latitudes of these stations observed, the difference of latitude will plainly be equal to the angle between the radii of the earth drawn from its center to the two stations, if the plumb lines with reference to which the latitudes are measured are in the directions of those radii. If, however, a great mass, like a mountain, stands between the two stations, the plumb bob south of it will be drawn toward the north, so that the latitude measured at that station will be less than the true latitude. The plumb bob north of the mountain will be drawn toward the south, so that the latitude measured at the northern station will be greater than the true latitude. If we know the true difference of latitude between the two stations, from a direct measurement of their distances apart and our knowledge of the size of the earth, we may determine, by the observations here indicated, the amount by which the plumb bob has been drawn aside by the attraction of the mountain, and hence may determine the relative values of the attraction of the mountain and of that of the earth. We may then determine the actual mass of the mountain from its dimensions and the average density of its parts, obtained from a study of specimens of the materials which compose it. From these data, assuming the law of gravitation, we may calculate the mass of the earth and hence its average density. By an experiment like the one here outlined, Maskelyne obtained for the mean density of the earth the value 4.7.

Another essentially similar method was employed by Carlini in 1824. He observed the oscillations of a pendulum at the bottom and at the top of a mountain. Assuming the law of gravitation, he calculated what the acceleration should be at the top of the mountain if its difference from that at the bottom depended solely on the change of level, and compared this calculated result with the acceleration observed at the top of the mountain. He found that the observed acceleration was greater than the calculated one. The difference he ascribed to the attraction of the mountain. By proceeding from this point as in Maskelyne's experiment, he obtained for the density of the earth the value 4.8.

The experiments of Cavendish and of von Jolly may also be

used in the determination of the earth's density. The results obtained in these experiments are much more consistent with one another that those obtained by methods in which the attraction of large parts of the earth is observed, and may be received with greater confidence. The value of the mean density of the earth obtained from them is about 5.64. The very careful experiments of Richarz and Krigar-Menzel give for the value of this constant the slightly smaller number 5.507 ± 0.009.

EXAMPLES, XI

1. *To show that the force exerted by a uniform spherical shell of gravitating matter, on a mass placed at any point within it, vanishes.*

Choose any point O within the spherical shell (Fig. 112), and with it as apex construct a double cone of small angular aperture ω. This cone will cut off sections of the surface at P_1 and P_2, which are equally inclined to the axis of the cone, so that their areas s_1 and s_2 conform to the proportion

$$s_1 : s_2 = OP_1^2 : OP_2^2.$$

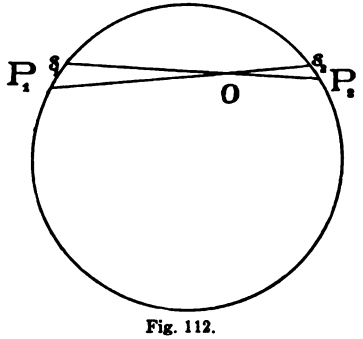


Fig. 112.

If σ represents the surface density, as it may be called, that is, the mass contained in unit area of the shell, the force with which a test mass μ at O is attracted toward P_1 is $k\mu\sigma s_1/OP_1^2$; and the force with which it is attracted toward P_2 is $k\mu\sigma s_2/OP_2^2$. The combined forces acting along the axis of the cone are therefore equal to $k\mu\sigma(s_1/OP_1^2 - s_2/OP_2^2)$, and from the proportion above, this sum equals zero. Since the whole sphere can be divided into similar portions by cones with their apices at O, it follows that the total force at O is zero.

2. *To show that the force exerted by a uniform spherical shell of gravitating matter, on a mass placed at any point outside of it, is proportional to the mass of the shell and inversely proportional to the square of the distance of the outside point from the center of the sphere.*

Join the center C of the spherical shell (Fig. 113) with the outside point O, and on the line CO find the point D, so placed that $OC \cdot CD = a^2$, using a to represent the radius of the sphere. (The point D is the foot of the perpendicular on the line CO from the point of tangency B of the tangent passing through O.) From O draw any line OP_1P_2 in the plane of the figure cutting the sphere in the points P_1 and P_2. Join C and D with P_1 and P_2. The triangles CDP_1 and OP_1C are similar, since they have the common angle OCP_1 and the sides about that angle proportional ($OC : CP_1 = CP_1 : CD$). In like manner the triangles CDP_2 and OP_2C are similar. In particular, the angles of these triangles at O, P_1, and P_2 are equal.

With O as apex describe a cone of small angular aperture about the line OP_1P_2 as axis. This cone will cut out areas s_1 and s_2 on the surface of the sphere which are equally inclined to the axis of the cone, so that $s_1 : s_2 = OP_1^2 : OP_2^2$. If

we represent the surface density by σ, the forces exerted on the test mass μ at O by the masses on these areas are $k\mu\sigma s_1/OP_1^2$ and $k\mu\sigma s_2/OP_2^2$; and from the proportion above, these two forces are equal, and the force exerted along the axis of the cone is $k2\mu\sigma s_1/OP_1^2$. The component of this force along the central line OC is $k2\mu\sigma s_1\cos\theta/OP_1^2$, if we use θ to designate the angle at O; and the component perpendicular to the central line OC is $k2\mu\sigma s_1\sin\theta/OP_1^2$. When the forces from all parts of the sphere are taken into account, this perpendicular component is counteracted by an equal and opposite one, as may be seen at once from symmetry; so that the only effective component of the force along OP_1P_2 is that along the central axis OC.

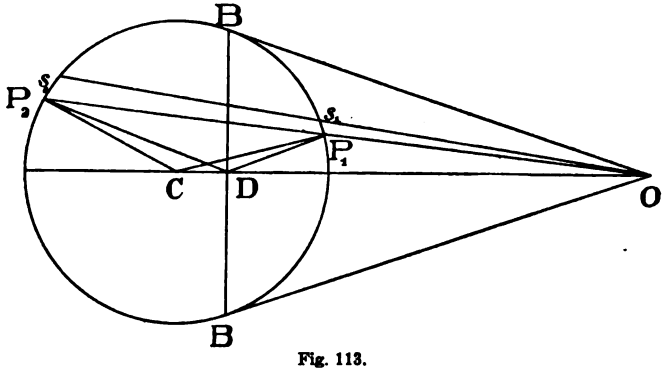


Fig. 113.

The whole sphere may be divided into areas similar to the areas s_1 and s_2, to which the same discussion will apply; and consequently the force on the mass μ at O due to the whole sphere is given by $\Sigma k2\mu\sigma s_1\cos\theta/OP_1^2$, taken over the part of the sphere lying nearer the point O than the plane passing through D perpendicular to OC.

We must now effect the summation indicated. To simplify the formulas we will designate the line OC by R, the line OP_1 by r, the line CP_1 by a, the line DP_1 by b. The projection of s_1 perpendicular to the line DP_1 is $s_1\cos\theta$, and the solid angle ω subtended by s_1 at D is therefore $\omega = s_1\cos\theta/b^2$. Using this in the typical term of the summation, it becomes $k2\mu\sigma\omega b^2/r^2$. From the similar triangles CDP_1 and OP_1C we have $b:a = r:R$, and using this in the typical term, it becomes $k2\mu\sigma\omega a^2/R^2$. Now all the factors in this term are constants except ω; and when the sum is taken to find the total force, the result is the product of these constant factors and the sum of all the solid angles subtended at D by the various areas of which s_1 is the type. This sum is equal to the angle subtended by the hemisphere of which D is the center, and which is limited by the plane through D perpendicular to OC, or to 2π. The total force, therefore, due to the sphere and acting along OC is $k4\pi a^2\sigma\mu/R^2$. But $4\pi a^2$ is the area of the shell, and $4\pi a^2\sigma$ is the mass M of the shell; so that the force due to the shell on the mass μ at O is $k\mu M/R^2$, and is directly proportional to the mass of the shell and inversely proportional to the square of the distance from the outside point to the center of the shell.

This theorem obviously applies to a solid sphere made up of spherical shells of uniform density.

3. *To find how Kepler's third law is modified when the motion of the planet is not about an attracting center, but about the sun, a central body which is also free to move.*

The center of mass of the two bodies remains stationary. Representing the masses of the sun and planet by m and μ respectively, we have the force on each given by $km\mu/r^2$, and the acceleration of the sun toward the stationary center of mass by $k\mu/r^2$. If we superpose on both bodies an acceleration equal and opposite to this, the sun will be stationary, and the planet will have the acceleration $k(m+\mu)/r^2$ with respect to it. This acceleration is to be substituted for km/r^2 in the proofs of Kepler's laws (§§ 140, 141). The motion of the planet will still be in an ellipse, one focus of which is situated in the sun, so that the planet will obey Kepler's second law. The formula connecting the periodic times and the semi-major axis will be modified by substituting $m+\mu$ for m, so that we have $m+\mu=\frac{\pi^2a^3\mu}{T^2}$. The ratio a^3/T^2 should therefore not be strictly constant for the different planets, as stated generally in Kepler's third law, but should vary with the mass of the planet.

The effect of the mass of the revolving body becomes appreciable when the body is large relatively to the central body. For instance, the mass of the moon is about 1/80 that of the earth, and should be taken into account when comparing the motion of the moon with that of an imaginary satellite near the earth's surface.

4. *Two large gravitating spheres A and B are placed so that the distance between their centers is R. The radius of one of the spheres is a. On the line joining its center with that of the other sphere and just outside its surface at C is placed a small body, and diametrically opposite, just outside the same sphere at D, is placed another small body* (Fig. 114). *Find the accelerations which these bodies will have toward the common center of mass, in the determination of which the small bodies may be neglected.*

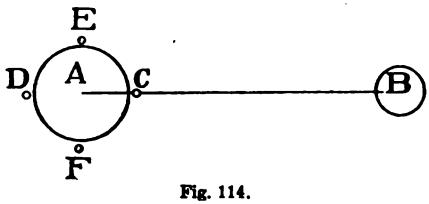


Fig. 114.

We represent the mass of the sphere A by E, that of the sphere B by M. Calling accelerations in the direction from A to B positive we have

$$acc_A = k\frac{M}{R^2}, \quad acc_B = -k\frac{E}{R^2},$$

$$acc_C = -\frac{kE}{a^2} + \frac{kM}{(R-a)^2},$$

$$acc_D = \frac{kE}{a^2} + \frac{kM}{(R+a)^2}.$$

The differences of the accelerations of C, A, D, are

$$acc_C - acc_A = -\frac{kE}{a^2} + \frac{kM}{(R-a)^2} - \frac{kM}{R^2},$$

$$acc_D - acc_A = \frac{kE}{a^2} + \frac{kM}{(R+a)^2} - \frac{kM}{R^2}.$$

If R is large relative to a we may neglect a^2 in comparison with R and obtain

$$\frac{kM}{(R-a)^2} = \frac{kM}{R^2 - 2aR} = \frac{kM}{R^2}\left(1 + \frac{2a}{R}\right);$$

and similarly

$$\frac{kM}{(R+a)^2} = \frac{kM}{R^2}\left(1 - \frac{2a}{R}\right).$$

In this case we have

$$acc_C - acc_A = -\frac{kE}{a^2} + \frac{2kMa}{R^3},$$

$$acc_D - acc_A = \frac{kE}{a^2} - \frac{2kMa}{R^3}.$$

If A is the earth and B the moon, we have $R = 60a$, and these formulas are permissible as a first approximation. In this case kE/a^2 directed toward the earth is the acceleration due to gravity, and the term $2kMa/R^3$ expresses the amount by which the acceleration of a body is changed when it is under the moon or on the opposite side of the earth from the moon. It will be seen that in both cases the acceleration toward the earth is diminished. The weight of bodies in these positions will therefore be diminished by the attraction of the moon.

At the quarter points E and F the acceleration toward the moon will be the same as that of the earth, and will not affect the weight of bodies placed there. If the earth were covered with a uniform layer of water, the weight of those parts under the moon and on the opposite side of the earth from the moon would be less than that of the parts at the quarter points, and to make equilibrium of liquid pressure throughout the mass there would have to be deeper water under the moon and on the opposite side of the earth from the moon than there is at the quarter points. This is a rough sketch of the equilibrium theory of the tides, which was worked out by Newton, and used by him as an additional proof of the validity of his theory of gravitation. As the earth revolves there will be high water as the place of observation passes under the moon, and after half a revolution there will be high water again, when the place of observation is on the opposite side of the earth from the moon. This principal feature of the tides is thus accounted for by the equilibrium theory.

If we call the mass of the earth 1 and its radius 1, and the mass of the moon 1/80 and its distance from the earth 60, the formula gives for the diminution of a body's weight due to the moon's attraction 1/8,640,000 parts of its weight.

The tides raised by the attraction of the moon are really waves of a length equal to the semi-circumference of the earth, and in water no deeper than the oceans such waves will travel freely with less velocity than the crest of the tide travels. The case then becomes one of forced vibration (Examples, V, 4), in which the period of the impressed force is less than that of the free vibration. The crest of the tide should therefore be in opposite phase with the moon, or at the quarter points when the equilibrium theory would expect it under the moon or on the opposite side of the earth from the moon. This is the dynamical theory of the tides.

The interruption of the tidal wave by the continents renders the phenomena of the tides very unlike those predicted by any elementary theory.

5. *How does the acceleration of gravity depend upon the elevation above the earth's surface?*

Let h represent the elevation and a the earth's radius; f the force of gravity at the station, and g its value at the base station; then $f : g = \frac{1}{(a+h)^2} : \frac{1}{a^2}$; and since h is very small compared to a we have

$$\frac{1}{(a+h)^2} = \frac{1}{a^2+2ah} = \frac{1}{a^2}\left(1 - \frac{2h}{a}\right);$$

so that

$$f = g\left(1 - \frac{2h}{a}\right).$$

6. *To calculate the work done when a unit mass of gravitating matter moves toward an attracting center through any distance.*

To simplify the formulas, we may express the force on a unit mass at the distance r from the attracting center by m/r^2. We wish to find the work done on the unit mass as it moves from X to A (Fig. 115) toward the center O. To do this we suppose the distance AX divided into many parts. We will call the distance OX, x; the distance OV, v; etc. As the unit mass moves through the distance $XV = x - v$, the average force on it is greater than m/x^2, and less than m/v^2. The work done is therefore greater than $\frac{m}{x^2}(x-v) = m\frac{v}{x}\left(\frac{1}{v} - \frac{1}{x}\right)$; and is less than $\frac{m}{v^2}(x-v) = m\frac{x}{v}\left(\frac{1}{v} - \frac{1}{x}\right)$. Using $W(X, V)$ to represent this work we may write

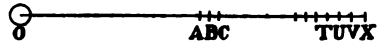


Fig. 115.

$$m\frac{v}{x}\left(\frac{1}{v} - \frac{1}{x}\right) < W(X, V) < m\frac{x}{v}\left(\frac{1}{v} - \frac{1}{x}\right);$$

$$m\frac{u}{v}\left(\frac{1}{u} - \frac{1}{v}\right) < W(V, U) < m\frac{v}{u}\left(\frac{1}{u} - \frac{1}{v}\right);$$

. . . . . .

$$m\frac{b}{c}\left(\frac{1}{b} - \frac{1}{c}\right) < W(C, B) < m\frac{c}{b}\left(\frac{1}{b} - \frac{1}{c}\right);$$

$$m\frac{a}{b}\left(\frac{1}{a} - \frac{1}{b}\right) < W(B, A) < m\frac{b}{a}\left(\frac{1}{a} - \frac{1}{b}\right).$$

The sum of these quantities in column will be less than, equal to, and greater than the work done in the movement from X to A. If we so proportion the lengths $a, b, \ldots v, x$ that the ratios $a/b, b/c, \ldots v/x$ are all equal, or if any one of them, say a/b, is less than any of the others, we shall have

$$m\frac{a}{b}\left(\frac{1}{a} - \frac{1}{x}\right) < W(X, A) < m\frac{b}{a}\left(\frac{1}{a} - \frac{1}{x}\right).$$

Now by making the divisions small enough, we may make the ratio a/b approach unity by increasing, and the ratio b/a approach unity by decreasing; and in the limit we shall have $W(X, A) = m\left(\frac{1}{a} - \frac{1}{x}\right)$.

If the point X is infinitely distant from O, we have $x = \infty$, and $W(\infty, A) = \frac{m}{a}$. The potential energy lost by the unit mass is measured by the work done on it, so that the excess of its potential energy at infinity over that which it has at the point A is m/a.

We shall subsequently define the *potential* at a point by the work which will be done on a unit mass by the force acting on it from the system (in this case,

the attracting center) as it moves from the point to infinity. The potential at the point A due to an attracting center at O is therefore $-m/a$.

7. *Show that the work done on a unit mass as it moves from one point to another under a force directed toward an attracting center is independent of the path.*

To do this, the path is cut up into elements, as in the previous example; the work done in each element is equal to the average force directed toward the center multiplied by the projection of the element upon the line of the force. The terms expressing it in each case therefore have the same form as those of the previous example, and the work done is given again by $W(X, A) = m\left(\frac{1}{a} - \frac{1}{x}\right)$, and is simply dependent on the distances of X and A from O, and not on their positions.

8. *To show that the force acting on a unit mass at any point from an attracting center is the rate at which the potential changes with the distance from that center with the sign changed.*

This is the reverse of the theorem in Example 6. The potential at any point whose distance from O is r is $V = -\frac{m}{r}$. Suppose r to increase by the small increment h, so that the potential becomes $V' = -\frac{m}{r+h}$. The change of potential $V' - V = -\left(\frac{m}{r+h} - \frac{m}{r}\right) = \frac{mh}{r^2 + hr}$. If h is very small we may neglect hr in comparison with r^2, and obtain $\frac{V' - V}{h} = \frac{m}{r^2}$. Now $\frac{V' - V}{h}$ is the rate at which the potential changes as r increases, and $-m/r^2$, or this rate of change of potential with the sign changed, is the force acting in the direction of the increase of r.

9. *To find the force on a unit mass at any point within a gravitating sphere of uniform density.*

From Example 1 it is evident that there will be no force exerted on the unit mass by that part of the sphere which lies outside the spherical surface whose radius is the distance r of the mass from the center. The force due to the part of the sphere inside that surface will be k (volume) (density)$/r^2$. Calling the density ρ, this becomes $\frac{4}{3}k\pi\rho r$. The force, and therefore the acceleration, is proportional to the distance of the unit mass from the center.

10. *To find the effect of a hill on the latitude of a station near it.* (*Maskelyne's experiment*, §148.)

We shall assume that the hill is a hemisphere. The force exerted by it on a body at its foot has a horizontal component and a vertical component. The vertical component will act so as to diminish slightly the weight of the body, but it is so small in comparison with the weight that the effect will not be noticed. The horizontal component will be at right angles to the weight. Calling the horizontal acceleration f, and using θ for the angle made by the plumb line with the true vertical, we shall have $f = g \tan \theta$. This angle θ can be determined by observations of the latitude.

The horizontal component of the attraction of the hemisphere is half that of the full sphere. If the radius of the hemisphere is r, and its density ρ', we have $f = k\frac{2}{3}\pi\rho' r$; while for the earth, whose radius is R and mean density ρ, we have $g = k\frac{4}{3}\pi\rho R$. Hence

$$\frac{f}{g} = \tan\theta = \frac{1}{2}\frac{\rho'}{\rho}\frac{r}{R}.$$

In Maskelyne's experiment, r was about 0.4 miles; ρ', about 2.8; θ, 6 seconds of arc. Taking $R = 4000$ miles, we get $\rho = 4.74$.

11. *To find the gravitational force due to a thick spherical shell at a point on its surface.*

Represent the outer and inner radii by R and R_1 respectively. The force due to the shell is the difference between the force due to a solid sphere of radius R, and that due to a sphere of radius R_1. If the density is ρ, we have therefore

$$f = k\frac{4}{3}\pi\rho\,\frac{R^3 - R_1^3}{R^2}.$$

If the thickness of the shell is d, we have $R_1 = R - d$, and if d is small compared with R, neglecting terms in d beyond the first,

$$f = k\frac{4}{3}\pi\rho \cdot 3d = k4\pi\rho d.$$

If σ represents what we may call the surface density, the mass of the shell under unit area, then $\sigma = \rho d$, and

$$f = k4\pi\sigma.$$

12. *To find the force at a point within a sphere of gravitating matter (Airy's experiment).*

We assume that the density ρ' is uniform in the shell whose inner radius is $R_1 = R - d$, when d is the depth of the point below the surface and R is the external radius. The mean density ρ of the sphere within the shell may be different from ρ'. The acceleration g at the outer surface due to both the shell and the inner sphere is

$$g = k\frac{4}{3}\pi\rho\,\frac{R_1^3}{R^2} + k4\pi\rho' d.$$

The shell exerts no force on a body within it (Example 1), so that the acceleration at the interior point is

$$f = k\frac{4}{3}\pi\rho\,\frac{R_1^3}{R_1^2}.$$

The ratio of these is $\dfrac{g}{f} = \dfrac{R_1^2}{R^2} + \dfrac{3\rho' d}{\rho R_1}$.

Substituting $R - d$ for R_1 and neglecting terms in higher powers of d than the first, this becomes

$$\frac{g}{f} = 1 - \frac{2d}{R} + \frac{3\rho'}{\rho}\frac{d}{R}.$$

Airy made comparisons of the value of g at the surface and f at the bottom of a mine in Cornwall. He found that a pendulum which beat seconds at the surface, or made 86,400 beats in a day, would make 86,402.24 beats in a day at the bottom of the mine. This gives

$$\frac{f-g}{f} = \frac{1}{19{,}286};$$ and from the formula we have $$\frac{f-g}{f} = \frac{2d}{R} - \frac{3\rho' d}{\rho R}.$$

The depth d was 1256 feet, the radius R may be taken as 3978 miles, and Airy assumed ρ' to be 2.5. With these data we obtain $\rho = 6.60$. This result is unduly great, no doubt because of an overestimate of the value of ρ', the mean density of the external spherical shell.

It is noteworthy that owing to the greater average density of the inner sphere, the acceleration at the bottom of the mine was greater than that at the surface.

CHAPTER VI

ELASTICITY

149. Cohesion. — When we try to break apart a piece of metal or wood, we encounter a very considerable resistance. This resistance is a manifestation of forces which act between the parts of the body. If we make the supposition that the matter of the body is so distributed as to occupy a considerable part of its volume, it may be shown by analysis that these forces cannot be due to the attraction of gravitation between the parts, the forces arising from gravitation being so small as to be negligible in comparison with them. The parts of the body cohere with one another, and the forces between them are called *forces of cohesion.* They may be considered attractions between the parts of the body.

When we attempt to compress the body into a smaller volume, we again encounter a resistance, which may be ascribed to a repulsion between the parts of the body. It is not certain that repulsive forces really exist. The effects ascribed to them may also be ascribed to the motions of the parts of the body, due to the heat of the body. We shall, however, adopt this hypothesis of repulsive forces provisionally. If we do so, we must consider the form and size of a body, in any condition in which its parts are relatively at rest, to be determined by a balance between the attractions and repulsions in every part of it

Because the structural elements of which a body is formed are called molecules, these forces between the parts of the body are often called *molecular forces.*

150. Elasticity. — If a metallic wire is fastened at one end and a weight hung on the other, the wire will stretch a little. The addition of another weight will stretch it further. When the weights are removed, provided the wire has not stretched beyond a certain limit, its length will become what it was at first. Phenomena essentially similar are exhibited by a rod when it is bent, a wire when it is twisted, a column on top of which a heavy weight is placed, or a mass of water or air confined in a vessel and subjected to pressure. The body is not displaced as a whole, but the forces which are applied to it change its shape or size, and

when they cease to act, the body returns more or less completely to its former condition. In each case of this sort the forces which counteract those applied to the body are evidently the forces of cohesion and of repulsion which have just been considered. The system of forces or pressures applied to the body may be called in general a *stress*. The change in shape or size or in both which the body undergoes may be called its deformation or *strain*.

As a result of common observation we may say that a stress applied to a body always produces a strain in it. If the body recovers its original condition, either in whole or in part, when the stress is removed, it is called an elastic body. The forces which it exerts against the applied stress are called elastic forces, and the *elasticity* of the body is the property which it possesses of exerting such forces. So far as our experiments go, there is no body which does not possess elasticity, if not with respect to all stresses, at least with respect to certain types of stress.

151. **Hooke's Law.** — In 1679 Hooke gave an account of experiments which he had tried on the stretching of wires by weights. He found that for wires of the same length and thickness, and of the same material, the elongations produced by the addition of different weights were proportional to the weights. He stated this result of his observations in the law, *Ut tensio, sic vis*, — The extension is proportional to the force.

This law is of far wider application than may be thought from the way in which it was derived. It applies to all cases of strains produced by stresses. If the stress is of a certain type, it will produce a strain of the same type. Thus, for example, a pull on a wire will lengthen it; a couple applied to one end of a wire, whose other end is fixed, will twist it; a pressure applied to a mass of water will compress it. In these cases, and in all others which might be cited, the law holds that the strain is proportional to the stress. This general law, to which all elastic bodies conform, we may call Hooke's law.

152. **Compression and Expansion.** — A pressure applied to a surface so as to be equal at every point on that surface, and normal to it, is called a hydrostatic pressure (§ 119). It may be either positive or negative. The pressure directed outward from the surface is taken as positive; that directed inward is negative. The bodies which we consider, in our present discussion of elasticity, are supposed to be homogeneous and isotropic; that is, they are alike in every part and possess similar properties, with respect to

elastic forces, in every direction. A hydrostatic pressure applied to such a body will change its volume without changing its shape. When the pressure is positive, the volume of the body increases, or the body undergoes expansion. When the pressure is negative, the body undergoes compression. In the case of compressions and expansions, Hooke's law may be stated by saying that the change of volume of the body is proportional to the change in the pressure applied to it.

The change in volume of a body, caused by a pressure, is manifestly proportional to the number of units of volume in the body. Bodies of different sorts exhibit characteristic changes of volume for the same changes of pressure. When the pressure changes by one unit, the decrease or increase of a unit volume will measure the compressibility or expansibility of the substance composing the body. This quantity, while it is the one most directly open to experimental examination, is not the quantity commonly employed to represent the characteristics of the body under pressure. We use instead the ratio of the increase in pressure to the change of volume of a unit volume. This quantity is called the *volume elasticity*, or often simply, the *elasticity*, of the substance composing the body.

Let the volume of a body be represented by v when under the pressure p. If the pressure is changed to p', the volume becomes v'. The change of volume is proportional to the change of pressure, and also to the volume v, so that we may write $k(v'-v)=v(p'-p)$, where k is a factor of proportion; and have

$$k=\frac{p'-p}{v'-v}v. \tag{73}$$

The factor k is the elasticity of the substance composing the body.

153. **Elasticity of Traction.** — When one end of a body is fixed, and a force is applied to the other end to lengthen it or to shorten it, the force applied is called a *traction*. The tractions commonly observed are pulls which lengthen the body, and the bodies used are long thin ones, like wires or rods. For such a body Hooke's law takes the form that the elongation of the body, that is, its increase in length, is proportional to the stretching force. When we examine wires of the same material, but of different lengths and cross sections, we find that their elongations are proportional to their lengths and inversely proportional to their cross sections.

It may easily be seen why this should be so, if we consider that the applied force produces a tension in every part of the wire, which is distributed as a nor-

mal pressure across the area in every cross section. This tension elongates each element of length similarly, and so the elongation of the whole wire is proportional to its length. The tension across each unit of area of a cross section is inversely as the area, and hence the elongation is inversely as that area.

Wires of equal lengths and cross sections, but of different materials, will exhibit different elongations when acted on by equal forces.

If we let e represent the elongation, F the force, l the length, S the cross section, and μ a factor of proportion, we may express these facts in the formula $\mu e = \frac{Fl}{S}$, and have

$$\mu = \frac{Fl}{eS}. \tag{74}$$

The factor μ is different for each substance, and is characteristic of it. It may evidently be described as the ratio of the applied force to the elongation which it will produce in a wire of unit length and of unit cross section. It is often described, on the supposition that the wire can be lengthened indefinitely without changing its properties, as the traction which will double the length of a wire of unit cross section. It is called the *modulus* of tractional elasticity, or often *Young's modulus*, having been introduced by Thomas Young in 1807 as a convenient characteristic number for elastic bodies undergoing traction.

154. Elasticity of Torsion. — When a wire clamped at one end is twisted by a couple applied to the other end, a pointer attached to it in the plane of the couple will turn through an angle which is proportional to the moment of couple. The wire is said to undergo *torsion*, and Hooke's law for the case of torsion is given by the statement that the amount of torsion is proportional to the moment of couple. Experiments on wires of different lengths and cross sections show that the amount of torsion is proportional to the length of the wire and inversely to the square of the cross section. It also depends on the material constituting the wire. The results of these experiments are collected in the formula

$$n = \frac{2\pi Cl}{\phi S^2}, \tag{75}$$

in which ϕ represents the amount of torsion, C the moment of couple, and $n/2\pi$ the factor of proportion. The quantity n, called the *modulus of rigidity*, or the *rigidity*, is different for different substances, and characteristic of them. It may evidently be described as the ratio of 2π times the applied couple to the amount of torsion

which it will produce in a wire of unit length and of unit cross section. Since the amount of torsion is measured in radians, and a torsion of 2π would be a complete rotation of the bottom layer of the wire through a full circle, it may also be described as the moment of couple which will completely rotate the bottom layer of a wire of unit length and of unit cross section.

155. Elasticity of Flexure. — If a straight rod or bar is clamped at one end, and if the other end is pulled aside by any force, the bar is bent or flexed. The amount of the *flexure*, determined by the distance through which the end of the bar moves, is proportional to the force applied to produce it. This is the statement of Hooke's law for this case. The amount of flexure depends also on the dimensions of the bar and on the material of which it is composed. The elastic coefficient or modulus characteristic of any particular substance in this case is the same as the modulus of tractional elasticity.

That this should be the case may be seen if we consider a diagram representing a bent rod (Fig. 116). The line running down the middle of the rod is not changed in length by bending, while those parallel to it above it are elongated and those below it shortened. As each of these lines reacts against the bending force in a way that depends upon its elasticity of traction, it is plain that their combined effect will be measured in terms of that elasticity.

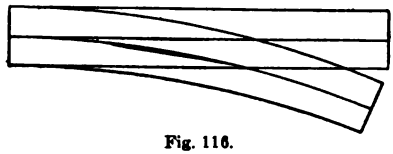


Fig. 116.

156. Work Done against Elastic Forces. — Let us consider a body which is strained and which conforms to Hooke's law. Then if x represents the amount of strain, which may be an elongation, a twist, a flexure, or in general, any strain, the force F which is required to produce it is proportional to it, and may be expressed by kx, where k is a factor of proportion. To find the work done when the body is strained from a condition in which its strain is x_0 to one in which it is x, we notice that since the force is proportional to the strain we may express the average force acting on the body as it is undergoing strain by $k(x+x_0)/2$, and since $x-x_0$ represents the amount of the strain, the work done by this average force is $k(x^2-x_0^2)/2$.

In the unstrained condition $x_0=0$, and the work required to impart the strain x is $kx^2/2$, and is proportional to the square of the strain. The elastic forces are conservative, and this expres-

sion is therefore also the measure of the potential energy of the strained body.

If the body is released and returns toward its unstrained condition, it will impart to any mass upon which it can act a simple harmonic motion. This appears from the fact that the force it exerts, and therefore the acceleration of the body moved by it, is proportional to the strain (§ 64). If the maximum strain is represented by a and any other strain by x, the diminution of potential energy, and therefore the increase in kinetic energy as the body moves in from the position of maximum strain, will be represented by $k(a^2 - x^2)/2$, so that we may write

$$\frac{1}{2}mv^2 = \frac{k}{2}(a^2 - x^2).$$

This relation is also characteristic of a simple harmonic motion (§ 63).

157. Homogeneous Strains. — In most of the theoretical discussions of the properties of elastic bodies we limit ourselves to the consideration of very small deformations. We suppose these deformations to be such that a line in the body which is straight before the deformation remains straight after the deformation. Strains of this sort are called *homogeneous strains.*

If we consider a sphere placed in the body which undergoes homogeneous strain, the strain which will keep all its radii straight lines will be composed of positive or negative elongations, in general of different magnitudes, along three mutually perpendicular axes. The sphere will thus become an ellipsoid, and the radii of the sphere will become the semidiameters of the ellipsoid. The principal axes of the ellipsoid are lines which are lengthened or shortened without rotation, and are called the axes of strain.

The elongations are by supposition very small. We shall designate the three coefficients of elongation by e_1, e_2, e_3. Then if we consider a cube with its sides of length a parallel to the axes of strain, it will strain into a parallelepiped, with sides $a(1+e_1)$, $a(1+e_2)$, $a(1+e_3)$. The volume of the parallelepiped will be $a^3(1+e_1+e_2+e_3)$, if we neglect the powers and products of the coefficients. The increase in the volume by the strain is $a^3(e_1+e_2+e_3)$, and the sum in parenthesis measures that increase. It is called the *coefficient of expansion.*

158. Fundamental Types of Strain. — A particular type of strain is obtained if $e_1 = e_2 = e_3$. It is that which has been called a *com-*

pression (§ 152). When a body undergoes strain of this type, the proportions of its parts remain unaltered, and the change is a change of volume without a change of shape.

Another particular type of strain is obtained if $e_1 = -e_2$, and $e_3 = 0$. This strain is called a *shear*. It is a strain for which the expansion is zero, and the body is lengthened along one axis, and proportionally shortened along a perpendicular axis, while no change occurs along the third perpendicular axis. A shear is a change of shape without a change of volume.

A shear may also be described as obtained by sliding contiguous planes in the body over each other.

159. Superposition of Strains. — If two successive elongations are impressed on a line in a body, its final elongation is equal to the sum of the particular elongations. For if l represents the length of the line, the first elongation makes it $l(1+e_1)$, and the second makes it $l(1+e_1)(1+e_2) = l(1+e_1+e_2)$, if we neglect the product e_1e_2.

An elongation can be obtained by the superposition of a compression and two shears. Let the two shears be those for which $e_1 = -e_2$, $e_3 = 0$, and $e_1 = -e_3$, $e_2 = 0$, and the compression one such that its value of e is equal to e_1. Then along the axis along which e_1 is measured (Fig. 117) there are three elongations, each equal to e_1. Along the axis along which e_2 is measured, the compression introduces an elongation $+e_1$, and the shear an elongation $-e_1$, so that there is no change in length along that axis. Similarly the compression and shear introduce equal and opposite elongations along the axis along which e_3 is measured, so that there is no change in length along that axis. The result of superposing a compression and two shears as described is therefore an elongation along one of the axes.

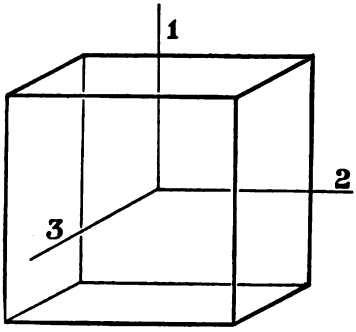


Fig. 117.

This being so, and any homogeneous strain being compounded of three elongations at right angles to one another, it is plain that any homogeneous strain may be made up of the proper combination of a compression and shears.

160. Fundamental Types of Stress. — Corresponding to the two fundamental types of strain, there are two types of stress, the

hydrostatic pressure and the *shearing stress*. The hydrostatic pressure (§ 119) is a pressure which is the same at all points of the surface to which it is applied and normal to the surface. The shearing stress (Fig. 118) is a pair of equal and opposite pressures applied tangentially to parallel surfaces in the body, combined with another pair of pressures similarly applied to two other surfaces in the body, at right angles to the first, in such a way that the system of four pressures introduces no moment of couple and serves merely to deform the body.

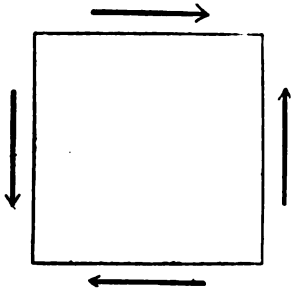


Fig. 118.

161. Fundamental Moduli of Elasticity. — Corresponding to the two fundamental types of stress and the two fundamental types of strain which they produce, we have the two fundamental moduli of elasticity, the *elasticity of volume* and the *rigidity*. Other moduli of elasticity, in particular Young's modulus, are functions of these fundamental moduli.

162. Failure of Hooke's Law. — No mention has been made, up to this point, of the fact that Hooke's law is not universally true. It is plainly not true for all possible strains, because, as we know, a body to which sufficient stress is applied may be permanently deformed or broken. When experiments are tried with stresses of different magnitudes, it is found that Hooke's law holds without perceptible error for small stresses, but fails when the stress exceeds a certain limit. The laws which have been stated in the foregoing sections must be interpreted in accordance with this statement.

The ideal body which obeys Hooke's law for all stresses, and which returns precisely to its original condition when the stress is removed, is called a *perfectly elastic body*. No real bodies are perfectly elastic with respect to both of the fundamental types of strain. Liquids and gases, possibly also solids, are perfectly elastic with respect to compression. No known body is perfectly elastic with respect to shear.

163. Elastic Fatigue. — The behavior of elastic solids under the action of forces, or which have been subjected to forces for a time and then relieved of them, depends in a curious and very complicated way upon the time during which the force acts and upon the time which elapses after the force is removed. The phenomena

exhibited by a body in these circumstances, and in others generally similar to them, are ascribed to what is called *elastic fatigue.*

164. Solids and Fluids. — Substances differ characteristically in the way in which they behave under stress. It is found by experiment that all known bodies may be grouped in one of two classes, according to the way in which they behave under shearing stress. The bodies in these two classes are called solids and fluids. A *solid* is a body which will offer a permanent resistance to a shearing stress; that is, it possesses a rigidity which depends upon permanent and conservative forces acting between its parts. The *fluid,* on the other hand, will not offer a permanent resistance to a shearing stress. If the shearing stress, no matter how small it is, is only applied long enough, the fluid will yield to it, will undergo a continual and increasing deformation, and will not tend to recover its original shape when the stress is removed. That is, the fluid possesses no true rigidity.

Fluids do not yield instantly and completely to a shearing stress. All fluids, even those whose parts move easiest among themselves, exhibit viscosity or internal friction, and in many fluids the viscosity is exceedingly great. When a shearing stress is applied to a fluid, the rate at which the fluid yields to it depends upon its viscosity, and the time taken to effect a perceptible deformation may be very great. Sealing wax is a good example of a body of this sort. If a stick of sealing wax is supported horizontally at its ends and a weight is placed on it, the stick will bend more and more with lapse of time. The wax, while thus conforming to the definition of a fluid, is so resistant to rapid motions of its parts that it is easily broken by a sudden blow.

EXAMPLES, XII

1. *To find the length of a wire, for which Young's modulus is μ, which is so long that its weight doubles the length of any short portion at its upper end, it being understood that its elastic properties are unchanged by such stretching.*

The density of the wire is ρ; its length and cross section L and S; its weight ρgLS; and hence the elongation of a length l at the top, to which this weight is applied, is $e = \frac{\rho gLSl}{\mu S}$. If the length l is doubled, $e = l$, and $L = \frac{\mu}{\rho g}$. This length is therefore a constant for the substance of which the wire is made, and is independent of the cross section of the wire.

2. *To find the work done when a wire of length L, cross section S, and with Young's modulus μ, is stretched by a weight W.*

When W is expressed in absolute units, the elongation e is given by $e = \frac{WL}{\mu S}$.

In the formulas of §156 we set $x = e$, $x_0 = 0$, $W = \frac{\mu S e}{L} = kx$, so that $k = \frac{\mu S}{L}$. The work is then given by $\frac{kx^2}{2} = \frac{W^2}{2} \cdot \frac{L}{\mu S}$.

3. *A rubber cord is stretched by a mass weighing W grams to double its length. One end of the cord is fixed on a smooth horizontal table, and the same mass pushed away from this fixed point, with the velocity v. How much will it stretch the cord?*

If e represents the elongation per unit length of the cord and l the whole length, we have the stretching force $= kel$, and if this force is $W = mg$, and the cord is doubled in length, so that $e = 1$, we get $k = mg/l$. The work done against the elastic force in any elongation el is $ke^2l^2/2$, and this is equal to the kinetic energy of the body, if the stretching is due to its motion, and reaches its maximum when the body loses all its velocity; so that $\frac{mv^2}{2} = \frac{ke^2l^2}{2}$; and using the value of k, we get

$$v^2 = e^2gl = e^2l^2\frac{g}{l}.$$

The elongation el is proportional to the velocity, and the factor of proportion is $\sqrt{\frac{g}{l}}$.

4. *A heavy body, whose moment of inertia, I, around a vertical axis can be calculated, hangs suspended by a cylindrical wire. By twisting the wire slightly, the body is turned out of its position of equilibrium, and is then released. Investigate its motion.*

The moment of couple exerted by the twisted wire is at each instant proportional to the angle through which its lowest plane has been turned. Representing the moment of couple by C, the angle just described by ϕ, we have $C = -k\phi$, where k is a factor of proportion. But the angular acceleration α of the moving body is given by $\alpha = \frac{C}{I} = -\frac{k\phi}{I}$. This equation is characteristic of an angular harmonic motion. The motion is harmonic with a period $T = 2\pi\sqrt{\frac{I}{k}}$.

From Equation 75 we may substitute for k in this equation, and obtain a formula for the value of n, the rigidity. Observations of the period of oscillation of the body, and of the other necessary data, enable us to determine the rigidity.

5. *Show how observations of oscillations due to torsion enable us to determine the moment of inertia K of an irregular body whose moment of inertia cannot be directly calculated.*

Using first a body like that of the last example, whose moment of inertia, I, is known, we determine T by observation. Then, attaching the body to be investigated to the wire, along with that already used, we determine the new period T' by observation. The wire remaining unchanged we have $T = 2\pi\sqrt{\frac{I}{k}}$, $T' = 2\pi\sqrt{\frac{I+K}{k}}$, and hence $K = I\frac{T'^2 - T^2}{T^2}$.

CHAPTER VII

CAPILLARITY

165. Capillarity. — If we dip one end of a glass tube in water and examine the water's surface around it and in it, we find that it is not everywhere level, as our hydrostatic theory asserts that it should be. Around the walls of the tube, both without and within, the water rises above the general level. If the diameter of the tube is small, the whole free surface within the tube rises above the general level, so that a column of water is lifted by it. This phenomenon is said to have been studied first by Leonardo da Vinci. The tubes in which it appears conspicuously have a very fine, or capillary bore. The general subject which deals with this phenomenon and with many others essentially like it, and due to the same general cause, is therefore named *capillarity*.

166. Forces of Cohesion. — The general laws of hydrostatics depend upon the principle that a liquid subject to the attraction of gravity will be in equilibrium only when its configuration is such that the action of gravity on it introduces no shearing stresses. Now gravity is not the only force which acts on the liquid. Its parts also exert forces of cohesion on one another, and true equilibrium will not be reached until the liquid assumes such a position that the combined action of the forces of cohesion and the weights of the parts of the liquid introduces no shearing stresses. We can explain all the phenomena which are treated under capillarity by taking these forces of cohesion into account. It is not necessary for us to know, and in fact we do not know, the way in which the cohesion depends upon the masses of the interacting parts and the distances between them. It is necessary, however, to assume this much, that the force of cohesion exerted by a small part or element of the body only acts on those elements which are in its immediate neighborhood. That is, we assume, as the law of the force of cohesion between elements, that the force between contiguous elements is very great, and diminishes very rapidly when they are separated, so as to become imperceptible, even when the distance between them is still very small. This general law of cohesive forces is illustrated by the behavior of an iron bar when we break it.

It requires a very great force to break it, but after it is broken, even though the two surfaces at the break are fitted together again with the utmost nicety, the two parts can be separated with no perceptible effort.

167. Surface Tension. — Because of the short distance within which an element of the liquid acts on its neighbors, those elements which lie below the surface by a depth equal to this distance, which we call the *range of action*, are in equilibrium under the action of their neighboring elements. It is only those elements which lie in or very near to the surface which are attracted unequally in different directions. These unequal attractions, acting on the elements of the surface film, will produce a peculiar condition in it. This will be the same for all parts of the surface, owing to the minuteness of the range of action, so long as the radius of curvature of the surface is not so small as to be of the same order of magnitude as the range of action. Thomas Young suggested that the special action of the cohesive forces in the surface film may be represented by supposing the film to be under tension, similar in general to that in a stretched membrane. This tension should be the same, in the case of any given liquid, for all parts of its surface. It is called the *surface tension*, and its numerical value, when determined for the surface of separation between any two bodies, is a characteristic number for those bodies. The position of the liquid column in the capillary tube, or any of the other phenomena ascribed to capillary action, are, on this view, due to equilibrium between the weights of the parts of the liquid and the forces due to the tensions acting in curved portions of the surface film.

The attempt which Young made to show that the existence of a surface tension can be deduced from the existence of forces of cohesion is generally admitted to have been unsuccessful, but the use of the concept of surface tension can be justified on other grounds. In order to do this, we shall use the method, introduced by Gauss, in which the capillary phenomena are ascribed to a peculiar potential energy resident in the surface film.

168. Surface Energy. — If we consider the action of the cohesive forces when the surface of a mass of liquid is enlarged by the movement of elements of the liquid from its interior into the surface film, it is plain that the potential energy of the liquid is increased as the surface is enlarged. For while each of the elements is moving out of the interior to take its place in the surface film, it is attracted inward by the forces of cohesion, and hence negative

work is done on it. The negative work thus done is equivalent to an increase in the potential energy of the liquid. The surface, therefore, possesses an energy peculiar to itself, which depends on the work done in enlarging it, and therefore on the magnitude of the cohesive forces and the thickness of the surface film. If the surface is free, the cohesive forces to be considered are those between the parts of the liquid alone; if it is bounded outwardly by another body, the forces between the elements of the liquid and the elements of that body must also be considered. The surface energy is proportional to the extent of surface, and is characteristic of the bodies separated by the surface.

By the aid of this concept of *surface energy* all the phenomena of capillarity may be explained by the use of the general principle that the potential energy of any material system tends to become the least possible.

169. Relation of Surface Tension to Surface Energy. — The admissibility of using the concept of surface tension may be shown by the aid of the surface energy. To do this, we consider a very simple and yet a general case, that of a film carried on a rectangular frame, one side of which is movable (Fig. 119). The two faces of the film are in every respect alike, and we need consider only one of them. We represent the surface energy in each unit of surface by E, and the length of the movable side by s. If this side slides in from its position at AB to a new position at $A'B'$, through the distance d, the area of the surface is decreased by the amount ds, and the potential energy Eds is expended. This expenditure of energy may be represented as due to work done by the film as it contracts. If we assume that a tension acts in the film, that can be represented as a force acting normally to any line in the film, and measured by the magnitude T of that force per unit of length of the line across which the tension acts, we may represent the force acting on the movable side s to draw it inward by Ts, and the work done by that force during the motion through the distance d by Tds. Setting this expression for the work equal to that for the corresponding potential energy expended we have $E = T$. We conclude that the same consequences can be deduced from the hypothesis of a surface tension acting in the film as from that of a surface energy resident in the film, and that consistent numerical

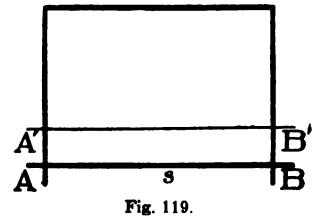


Fig. 119.

results will be obtained from these hypotheses if the numerical value of the surface tension is the same as that of the surface energy.

170. Pressure under a Curved Surface. — When a portion of the liquid surface is curved, the tension in the surface produces a normal pressure under it, which depends upon the curvature. It is to this pressure that the peculiar modifications of the ordinary laws of hydrostatics are due, which become apparent when the observations are made with sufficient minuteness to notice the phenomena of capillarity.

To show how this normal pressure depends on the curvature, we may first consider a surface of simple curvature, like a portion of a cylindrical surface. Such a surface exists on water elevated by capillary action between two parallel glass plates. Consider a very small rectangular area in this surface, two of whose sides are straight and parallel with the axis of the cylinder. The tensions acting across the curved sides, and therefore parallel with the axis, are exactly opposite in direction, and therefore act only to stretch the surface of the area, and have no component normal to the surface. The tensions acting across the other pair of sides are not exactly opposite in direction, and have a normal component. Let s and s' represent the lengths of the sides of the rectangular area, the side s being one of the curved sides. The radius of curvature of that side (Fig. 120), designated by ρ, is connected with the arc s, and the angle ϕ between the extreme radii drawn from the center of curvature C to the ends of the arc, by the formula $s = \rho\phi$. Now the tensions apply forces to the edges s' each equal to Ts', and since these forces are tangential to the surface at the extremities of the arc s, there is an angle between their directions equal to ϕ. They will therefore give rise to a normal resultant force (Fig. 121) equal to $2Ts'\sin(\phi/2)$, or since ϕ is very small, to $Ts'\phi$, or to Tss'/ρ. This normal force is a force pulling the area ss' inwards and gives rise to a pressure, or force per unit of area, under the surface, equal to T/ρ.

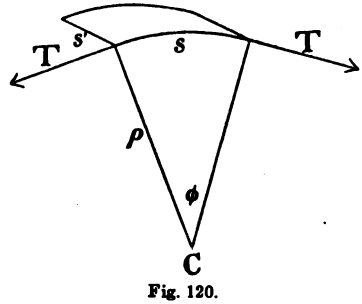


Fig. 120.

Fig. 121.

If the surface is one in which both the sides s and s' are curved, a similar argument will show that the pressure under such a surface

is given by $T\left(\frac{1}{\rho}+\frac{1}{\rho'}\right)$. From Euler's theorem we know that this sum $\left(\frac{1}{\rho}+\frac{1}{\rho'}\right)$ is constant at a chosen point on the surface for any position of the two rectangular plane sections for which the radii of curvature ρ and ρ' are measured, and hence we conclude that this expression represents generally the pressure under a curved surface due to the tension in it, and is independent of the particular orientation of the rectangular area with which it has been deduced.

Young obtained this result from his hypothesis that the tension in a liquid surface is similar to that in a stretched membrane. It was obtained almost contemporaneously by Laplace, by a method which took account of the action of the cohesive forces directly, and the formula for the pressure is commonly called Laplace's formula.

171. **Plateau's Experiments.** — If a limited portion of liquid is so situated that its equilibrium does not depend upon its weight, the form which the liquid will assume when in equilibrium will depend upon the forces of cohesion alone. In order to examine the behavior of a liquid in these circumstances, Plateau mixed a quantity of alcohol and water, adjusting the proportions of the two ingredients until the density of the mixture was the same as that of olive oil. When masses of olive oil were introduced into this mixture, they had no tendency to rise or sink, their weights therefore did not effect their equilibrium, and the cohesive forces were free to act alone. Each of the masses of oil assumed a spherical form, provided they were entirely free in the supporting mixture. This result is consistent with Laplace's formula; for equilibrium can exist in such a mass only when the pressure is the same in it everywhere, and in order that the pressure should be the same, the curvature of the surface film should be the same everywhere. This is true for the sphere, and for the sphere only among the possible finite and unbounded surfaces. Looking at the same matter from the point of view of surface energy, the form which the mass assumed was that having the least potential energy and therefore the least surface. This is well known to be the sphere.

When the mass of oil was not free, but suspended on a wire framework, its surface assumed different forms. Whatever was the form of these surfaces, they were such that tensions in them of equal value gave rise to equal pressures within the liquid mass. We see from Laplace's formula that this condition is attained

when the curvature of the surfaces is such that the sum of the reciprocals of the principal radii of curvature is the same at every point on them.

For example, a mass of the oil suspended on a circular ring (Fig. 122) formed a lens-shaped figure, with its two faces similar parts of equal spheres. A mass suspended between two parallel and equal rings concentric to the same line, had two spherical caps outside the planes of the rings and a portion with a variously curved surface between them. When the rings were properly placed, the surface between them became a right cylinder, the radius of which was half the radius of either of the spherical caps at the ends (Fig. 123).

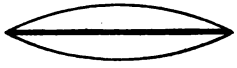


Fig. 122.

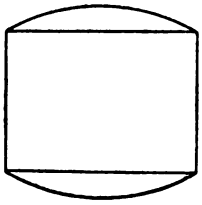


Fig. 123.

Plateau obtained similar surfaces by the use of films of soapy water. The weight of the film is so small a force, in comparison with the forces introduced by the surface tension, that the figure due to the surface tension is scarcely distorted at all by the weight of the film. When the film has no boundaries, it becomes a sphere containing air under higher pressure than that outside it. Since both the inside and outside surfaces of the film contribute equally to the pressure, it appears at once from Laplace's formula that the excess of pressure is equal to $4T/\rho$. When the continuity of the film is interrupted by boundaries, such as the parts of a wire frame, but in such a way that the film forms a surface enclosing air, the shapes of the various portions into which the film is divided by the boundaries are such that $T(1/\rho + 1/\rho')$ is the same for every point on them, and equal to the excess of pressure within the surface over that outside it.

When the film is carried on a frame or boundary, so that it does not enclose a volume, the pressure is the same on both its faces, and the shape is such that $1/\rho + 1/\rho'$ is everywhere zero. The simplest example of this case is that in which the boundary lies in a plane. The film is then everywhere plane, as may be seen also directly from the condition of minimum energy. If the boundary does not lie in a plane, the principal radii of curvature at a point on the surface are generally equal and oppositely directed. A surface whose principal radii of curvature are thus related is called a ruled surface. It is a property of such a surface that its area is less than that of any other surface limited by the same boundary. This geometrical property is consistent with the condition that the potential energy of the film is a minimum.

172. Meeting of Three Fluids. — When three fluids which do not mix meet along a line, the three surfaces which separate the fluids make with one another definite angles which depend upon the surface tensions in the surfaces. The surface tensions acting across any portion of the line of contact give rise to three forces proportional to the surface tensions, and the angles between the lines of these forces are determined by the parallelogram law. If we designate the three fluids by a, b, c, the tensions in the surfaces separating the fluids, two and two, by T_{ab}, T_{ac}, T_{bc}, and the angles between the surfaces by α, β, γ (Fig. 124), then by Lami's theorem (Examples, III, 2) we have for equilibrium

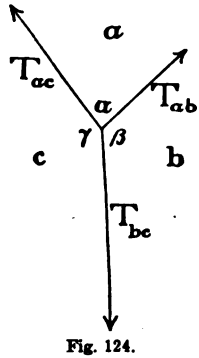


Fig. 124.

$$\frac{T_{bc}}{\sin\alpha} = \frac{T_{ac}}{\sin\beta} = \frac{T_{ab}}{\sin\gamma}.$$

These angles are finite when the tensions lie within certain limits of relative value, but when one of the tensions exceeds the sum of the other two, equilibrium cannot exist, and the line of contact is drawn in the direction of the greatest tension, so that one of the fluids, the one bounded by the surfaces in which the tensions are less, is extended over the surface separating the other two fluids.

For example, the tension in the surface separating water and its vapor is greater than the sum of the tensions in the surfaces separating oil and water, and oil and water vapor; so that a drop of oil placed on water will be drawn out into a thin sheet over the surface of the water. Since the surface energies in these surfaces are equal to the surface tensions, the condition that the potential energy tends to a minimum value is illustrated by the spreading out of the oil in such a way as to diminish the surface in which the surface energy is the greatest.

173. Angle of Contact. — When one of the three bodies in contact is a solid, the surface separating the two fluids makes a definite angle with the solid at the line of contact. This law was announced by Young as a fact of observation, but it follows immediately from the condition of equilibrium of the three surface tensions which are acting across the line of contact. The case of most frequent occurrence is that of a liquid standing under its own vapor, or under air, and in contact with a solid wall. Let us designate the solid by a, the liquid by b, its vapor or the air by c, the tensions in the surfaces separating these bodies by T_{ab}, T_{bc}, T_{ac} as before, and the angle of contact which the liquid-air sur-

face makes with the liquid-solid surface by α. Then (Fig. 125) since the line of contact can only move along the solid and not away from it, there will be equilibrium if the sum of the components of the three tensions tangential to the solid is equal to zero. This condition therefore gives $T_{ac} = T_{ab} + T_{bc}\cos\alpha$, and hence

$$\cos\alpha = \frac{T_{ac} - T_{ab}}{T_{bc}}, \qquad (76)$$

and α is a constant for the given solid and liquid. If T_{ac} is greater than T_{ab}, $\cos\alpha$ is positive and α is an acute angle. If T_{ac} is greater than $T_{ab} + T_{bc}$ the formula yields the impossible result that $\cos\alpha$ is greater than unity. In this case there is no equilibrium of the line of contact, but the liquid spreads out over the solid, and wets it, so that the contact angle between the liquid and solid becomes zero. An example of this is furnished by water in contact with a clean glass surface. If T_{ac} is less than T_{ab}, $\cos\alpha$ is negative and α is an obtuse angle. Such an angle may be observed when mercury is in contact with glass.

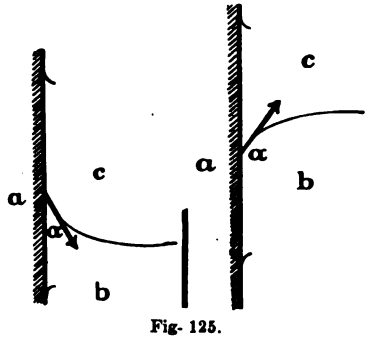


Fig. 125.

174. Equilibrium in Capillary Tubes. — If one end of a fine tube is immersed in a liquid which wets it, the liquid will rise in the tube. The height at which it stands when equilibrium is attained is directly proportional to the surface tension of the liquid, and inversely proportional to the radius of the tube. This statement is commonly referred to as *Jurin's law*.

To demonstrate this law we notice that the liquid surface at the line of contact makes an acute angle with the wall of the tube, and that the free surface of the liquid in the tube is consequently curved so as to be concave upwards. The surface tension in the curved surface tends to flatten it and thus to lift the liquid beneath it, and an upward movement of the liquid will occur and continue until the weight of the column lifted is in equilibrium with the vertical component of the force due to the surface tension of the liquid acting at the line of contact. Let us designate the surface tension of the liquid by T, the density of the liquid by d, the height of the column by h, the radius of the tube by r, and the contact angle by α. The component $T\cos\alpha$ of the tension will act upwards at every point

on the circumference of the surface, so that the lifting force is $2\pi rT\cos\alpha$. The weight of the column is approximately $gd\pi r^2h$; and equilibrium will occur when

$$2\pi rT\cos\alpha = gd\pi r^2h.$$

From this equation we get

$$h = \frac{2T\cos\alpha}{gdr}, \tag{77}$$

thus demonstrating Jurin's law, and at the same time obtaining a relation by means of which the surface tension may be determined. If the liquid really wets the tube, $\cos\alpha = 1$, and an observation of the height of the column and of the diameter of the tube will furnish the quantities which are needed in the calculation of the surface tension.

If the liquid does not wet the tube, so that the contact angle is obtuse, the top of the column in the tube is depressed below the general level. The same formula applies in this case, with $\cos\alpha$ and therefore h negative.

When the radius of the tube is small and the contact angle zero, the surface of the liquid is approximately a hemisphere, with its radius equal to that of the tube. This can be seen from Laplace's formula, for the height of the liquid column is approximately the same for all points of the curved surface, and therefore $1/\rho + 1/\rho'$ is the same for every point. Thus the liquid surface is a part of a sphere, and since the surface is tangent to the wall of the tube its shape is hemispherical and $\rho = r$. The portion of the liquid immediately surrounding the curved free surface is called the meniscus.

If a small portion of liquid is contained in a long capillary tube, whose bore is not cylindrical but conical, it will move toward the smaller end of the tube. The liquid is terminated by two hemispherical surfaces, of which the one nearer the smaller end of the tube has the smaller radius. By Laplace's formula, therefore, the negative pressure under that surface is greater than that under the other surface, and the liquid will move in response to it, unless equilibrium is established by tilting the tube so that the smaller end is raised until the weight of the liquid compensates for the greater negative pressure at that end.

If the column of liquid is in a cylindrical tube it will move toward the end in which the surface tension is greater. We may diminish the surface tension of one end of the column by raising its temperature, or by covering it with a film of liquid of lower surface tension.

175. Bubbles and Drops. — A bubble is a mass of air surrounded by a liquid. The shape of the liquid surface is determined by the condition that the pressure at every point, due to the surrounding liquid, in addition to the pressure due to the surface tension, shall be equal to the pressure of the air in the bubble. The bubble of typical form is one made by introducing air under a flat glass plate

held horizontally in the liquid. The plate keeps it from rising, and it assumes a condition of equilibrium determined by the balance of pressures already described. When the bubble is very minute, its form is approximately spherical. As more air is introduced it widens out and increases in height. Its height depends upon the surface tension of the liquid, and tends toward a limiting value, as the volume of the bubble increases, of not more than a few millimeters in any known case. Any horizontal cross section of the bubble is circular. Its under surface is flat, when the bubble exceeds a certain size, and over that flat surface the pressure of the air within the bubble is equal to the pressure of the liquid column whose height is the height of the bubble. For points on the side of the bubble, at which the surface is curved, the pressure which maintains equilibrium with that within the bubble, is partly that due to the depth of the point and partly that due to surface tension in the curved surface.

A drop is simply an inverted bubble, a mass of liquid resting on a support and surrounded by air. A drop of mercury on a flat surface is a typical form. With evident modifications, what has been said about the bubble applies equally well to the drop. Drops of water, or of other liquids which wet solids, are usually limited in respect of the area of the base on which they stand either by impurities on the surface of the solid, which prevent its being wetted, or by abrupt changes in the shape of the solid. The theory of the equilibrium of such drops is essentially the same as that already stated, but special cases sometimes need careful study in order to perceive that they are not inconsistent with the general principles already laid down.

176. Surface Viscosity. — The parts of some liquids move with much less freedom among themselves than those of others. The forces with which they resist the force which moves them are non-conservative forces and do not tend to restore the parts which have been displaced to their original position. They are ascribed to the mutual forces between the parts, coupled with the mobility of those parts. The cause of this action is called internal friction or viscosity.

In many liquids, probably in all of them, the viscosity in the surface film is different from that in the general mass. Its value is often very much greater than that in the mass. The value of a liquid, as a means of illustrating the principles of capillarity, often depends very much on the fact that it possesses a considerable

surface viscosity. For example, the surface tension of water is greater than that of any other ordinary liquid, and pure water would therefore seem to be the best liquid that could be used, in experimenting on capillarity. On the other hand, the surface viscosity of water is not great, and the parts of a film of pure water will move so freely under their own weight that the film very soon becomes somewhere thin enough to break. The addition of soap to the water, so as to make a strong soap solution, diminishes the surface tension, but very considerably increases the surface viscosity, so that a film of soap solution, whose parts do not move freely, will remain unbroken for so long a time that its form and peculiarities may be studied.

EXAMPLES, XIII

1. *Find the formula for the height to which a liquid which wets glass will rise between two parallel glass plates at the distance r apart.*

The surface tension acts vertically along the lines of contact of the liquid with the glass. Representing the density of the liquid by d, we get for the height h, $h = \frac{2T}{dgr}$. Compare with the formula for the rise in a capillary tube, § 174.

2. *Two vertical glass plates touch along one vertical edge and make a slight angle ϕ with each other. Show that the curve of the surface of liquid that rises between them is an equilateral hyperbola.*

The distance apart of the plates, at a distance x from the edges in contact, is proportional to that distance and may be set equal to ϕx. Using the result of the last example, the height of the liquid at that point is $h = \frac{2T}{dg\phi x}$; so that $hx = \frac{2T}{dg\phi}$. Now all the factors on the right are constants, so that hx = constant; and this is the formula of the equilateral hyperbola, of which the asymptotes are the line of the edges in contact, and a line at the level of the free surface.

3. *To prove Jurin's law* (§ 174) *by means of Laplace's formula.*

The surface of the liquid in the tube is a portion of a sphere. If the radius of the tube is r and the angle of contact is α, the radius of curvature ρ of that surface is $\rho = \frac{r}{\cos\alpha}$; and hence the upward pressure under that surface is equal to $\frac{2T}{\rho} = \frac{2T\cos\alpha}{r}$. This is equal to the pressure due to the height of the column, so that $\frac{2T\cos\alpha}{r} = dgh$, or $h = \frac{2T\cos\alpha}{dgr}$.

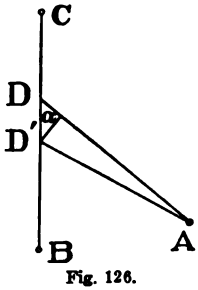


Fig. 126.

4. *To show that the condition that the surface energy shall be a minimum accounts for the angle of contact between a liquid surface and a solid.*

Represent the surface of a liquid in contact with a plane solid as in Fig. 126. Select points A, B, C in the surfaces, from which the respective distances a, b, c are measured to the line of

contact D. Call the surface energies in the surfaces T_a, T_b, T_c respectively. Then the energy in a strip of unit width in the surface AD is given by $T_a a$, and the total energy in the three strips in the three surfaces is $E = T_a a + T_b b + T_c c$. If the point D is slightly displaced by the distance d along the surface of the solid to D', the energy E' in the surfaces thus formed will be

$$E' = T_a(a - d\cos\alpha) + T_b(b - d) + T_c(c + d).$$

If the energy is a minimum in the original position of the surfaces, it will not be changed by this displacement, so that $E = E'$, from which we get at once

$$\cos\alpha = \frac{T_c - T_b}{T_a}.$$

CHAPTER VIII

GASES

177. Nature of a Gas. — The experiment of Torricelli with the barometer (§ 128), showed that it was possible to obtain a region in space that was to all appearance void or empty of matter. This region above the mercury of the barometric column was called the Torricellian vacuum.

It occurred to von Guericke (1602–1686) that it might be possible to produce a similar vacuum by withdrawing air from a closed vessel by means of a pump. After several unsuccessful attempts he at last succeeded in doing this, if not so completely as in the Torricellian vacuum, at least sufficiently well to enable him to study the behavior of bodies in a place void of air. One of his experiments directed attention to a peculiar property of air, by which it is distinguished from the other bodies with which the physicists of his time were familiar. He showed that if a partially collapsed bladder, tied tightly at the neck, so that the air in it could not escape, was placed in the vessel or receiver of his air pump, and the air removed from around it, the air in it would swell out until the bladder seemed full. Thus the air in the bladder appeared to expand in all directions. Its behavior in this respect is in marked contrast to that of a solid or of a liquid placed in the receiver, neither of which will expand in all directions when the air is removed. A body which exhibits this property of expansion in all directions, or, as it is sufficient to say, of exerting pressure in all directions on the walls of any closed vessel containing it, is called a gas.

178. The Air as a Fluid. — A limited portion of air has weight. This fact was recognized by Galileo, who attempted to prove it, without complete success, by first weighing a glass vessel full of air, and weighing it again after some of the air had been expelled by heating it. Aristotle failed in an attempt to prove the same thing. He weighed a bladder full of air, and then, after forcing the air out of the bladder, weighed it again. Of course he found no difference between the two weights, because the air which was outside the bladder in the second trial took the place of, and

weighed as much as, the air which was within it in the first trial. The experiment cannot succeed unless the walls of the vessel are rigid, and unless the air can be removed from its interior. By employing an experiment of this sort, von Guericke established the fact that air has weight.

This being so, and air being a fluid, the pressure relations of which are essentially the same as those of a liquid, it follows that, with certain modifications, hydrostatic theorems apply to air as well as to liquids. In particular Archimedes' principle applies to both classes of bodies. A body weighs less in air than it would weigh in vacuum, by an amount equal to the weight of the air which it displaces. In very exact weighing, it is therefore necessary to take account of the volumes of the bodies which are weighed, and of the weights used, in order to obtain the weights of the bodies in vacuum.

179. Gases. — For many years atmospheric air was the only gas known. In 1755 Black discovered the gas called carbonic acid or carbon dioxide. When attention was thus directed to the possible existence of other gases, the discovery of hydrogen and of many other gases was soon made; and it was also recognized that the vapors formed by the evaporation of volatile liquids were in most essential respects similar to gases. Thus to the two classes of bodies which had long been recognized, solids and liquids, there was added a third class, gases.

For our present purposes, we may define a *solid* as a body which retains its form and volume unchanged under the action of its own internal forces. A *liquid* is a body which retains its volume unchanged under the action of its own internal forces, but yields to shearing stress, so that it cannot retain a definite form, except when placed in some receptacle whose walls exert pressure on it of such a sort as to annul the shearing stresses in it. A *gas* is a body whose internal forces do not constrain it to assume any definite volume or form. It expands in all directions, so that it cannot be confined as a liquid can, in a vessel open at the top. When confined in a closed vessel, it exerts pressure upon every part of the surface which envelops it.

180. Boyle's Law. — The experiments of von Guericke with the air pump were repeated and amplified by Robert Boyle (1627–1691). His attention was thus attracted to the relation between the volume of a limited portion of air and the pressure upon it, and it occurred to him that it might be worth while to obtain a set

of values of the volumes and of the corresponding pressures. He did this by isolating a small quantity of air in the short limb of a U-shaped tube (Fig. 127) by means of a quantity of mercury in the bend of the tube. When this mercury was so adjusted that its ends stood at the same level in both limbs of the tube, the pressure on the enclosed air was that of the atmosphere. When more mercury was poured into the long limb of the tube the surface in the short limb rose and the air was compressed. The pressure on it was that of the atmosphere, increased by the pressure of a mercury column whose height was the difference of level between the surfaces of the mercury in the two limbs of the tube. By means of a number of measurements of this sort Boyle showed that the volume of the air in the tube varied inversely with the pressure upon it. Representing the initial and final volumes by v' and v, and the initial and final pressures by p' and p, Boyle's law in this form is expressed by the proportion

Fig. 127.

$$p : p' = v' : v. \tag{78}$$

To illustrate this statement we shall consider the air confined in the tube under atmospheric pressure. We may take the volume of the air in this condition as unit volume, and the pressure of one atmosphere as unit pressure. If mercury is then poured in until the pressure becomes two atmospheres, the volume of the air is reduced to one-half its original volume. If the pressure is made three atmospheres, the volume of the air is one-third its original volume, and so on. For each of these cases the product of the volume and the corresponding pressure is equal to 1. This product might have had any numerical value we pleased to give it, depending upon our choice of the volume taken as unit volume and the pressure taken as unit pressure. But whatever that value was, the experiments show that the product of the volume and the corresponding pressure will always be the same. We may therefore express Boyle's law in another and more convenient form, by saying that the product of the volume of a given mass of air and the pressure upon it is constant. In this form Boyle's law is expressed by the equation

$$pv = p'v'. \tag{79}$$

By the use of Boyle's law we may deduce the voluminal elasticity of air, and thus put the law in still another form. We consider the case in which the pressure p is very little different from the initial pressure p'. We have $p = \frac{p'v'}{v}$, and can form the difference of pressure

$$p - p' = \frac{p'v'}{v} - p' = p' \cdot \frac{v' - v}{v} = -p' \frac{v - v'}{v}.$$

Now $p - p'$ is the increase of pressure, and $(v - v')/v$ is the increase of volume

(necessarily negative) per unit volume, so that when p and p' are nearly equal, $v\frac{p-p'}{v-v'} = -p$ is by definition the modulus of voluminal elasticity. We may therefore express Boyle's law by saying that the voluminal elasticity of air is equal to the pressure.

Boyle's law was subsequently shown to hold approximately, not only for air, but also for all gases. If equal volumes of different gases are taken under equal pressures, the volumes of all of them will change, when the pressures are changed, according to the same law. The law holds not only for the case in which the pressure is continually increased, but also for that in which the pressure is decreased. Gases, therefore, differ from solids and liquids in that they all possess the same voluminal elasticity.

181. Gay-Lussac's Law. — Gases expand when heated. This fact was known to Galileo, who used it in the construction of the first thermometer. Owing to various causes, the law of this expansion was not discovered until many years after Galileo's time. In 1809 Gay-Lussac showed that all gases expand at the same rate as their temperatures rise (§ 260). That is, when the temperatures of different gases are raised from that of the melting point of ice to that of the boiling point of water, their volumes all increase by the same proportionate amount. Gases differ from solids and liquids in that they all possess the same coefficient of expansion with rise of temperature.

This law was discovered, though not published, by the chemist Charles. It is therefore sometimes known as Charles's law.

182. Law of Combining Volumes. — As the result of his experiments on the chemical combination of gases, Gay-Lussac showed that, in the case of complete chemical combination, the volumes of the gases which combine and the volume of the resultant product, if it also is a gas, are in a simple numerical relation to one another. For example, two volumes of hydrogen will combine completely with one volume of oxygen, and as the result of that combination, two volumes of water vapor are obtained. The volumes are of course measured at the same pressure and temperature. The law thus stated is called the *law of combining volumes.*

183. Avogadro's Law. — An explanation of the law of combining volumes was given in 1811 by Avogadro, and in 1814 by Ampère.

It had for some time been believed that bodies do not consist of matter continuously distributed, but that the matter of a body is collected in separate particles. Newton described these par-

ticles in the following words: "It seems probable to me that God in the beginning formed matter in solid, massy, hard, impenetrable, movable particles, of such sizes and figures and with such other properties and in such proportion to space as most conduced to the end for which He formed them; and that these primitive particles, being solids, are incomparably harder than any porous bodies compounded of them, even so very hard as never to wear or break in pieces; no ordinary power being able to divide what God Himself made one in the first creation. . . . It seems to me, farther, that these particles have not only a *vis inertiae*, accompanied with such passive laws of motion as naturally result from that force, but also that they are moved by certain active principles." The particles of gases were supposed to be of this sort.

Avogadro perceived that the law of combining volumes could not be explained by assuming the particles of the gas to be such particles as these. He therefore assumed that the particles which compose the gas, and which give to it its characteristic physical properties, are combinations of two or more elementary particles. When these elementary particles are of the same kind, the substance made up of them is a chemical element. When they are of different kinds, the substance is a compound. Chemical combination then involves a breaking up of the groups of each of the elements, and a combination of the particles composing these groups with each other to form new groups. The elementary particles are called *atoms*, the groups composed of them, *molecules*. These names were first given by Ampère.

Avogadro announced the law that equal volumes of different gases, under the same conditions of pressure and temperature, contain equal numbers of molecules. With this law as a foundation he was able to explain the law of combining volumes.

Without going into a complete discussion of the matter, we may illustrate the consistency of Avogadro's hypothesis with the law of combining volumes by one or two examples. In the case of the combination of oxygen and hydrogen to form water vapor, we assume that the molecules of oxygen and hydrogen are made up of two atoms each. According to Avogadro's hypothesis there are twice as many molecules of hydrogen as of oxygen in the volumes which give complete combination. When combination occurs, one atom of oxygen combines with two atoms of hydrogen to form a molecule of water vapor, so that there result twice as many water molecules as there were oxygen molecules originally, and the observed result that the volume of the water vapor is twice that of the oxygen is consistent with Avogadro's hypothesis. When equal volumes of hydrogen and chlorine combine to form twice the volume of the

vapor of hydrochloric acid, the new molecules contain one atom each of the two constituents of the combination. There is therefore a molecule formed for each atom of hydrogen or of chlorine, and the resulting volume contains twice as many molecules as there were in either of the two constituent gases; and the observed result that the volume is twice that of the hydrogen or of the chlorine is consistent with Avogadro's hypothesis.

No such law as this can be shown to hold for solids and liquids. Gases, therefore, differ from solids and liquids in that equal volumes of them, under similar conditions, contain equal numbers of molecules.

184. Molecular Theory of Matter. — This theory of the constitution of gases, which has been forced upon us by the study of their chemical properties, has been extended so as to afford a complete explanation of all the chemical reactions between bodies. We therefore conclude that a homogeneous body is composed of similar molecules, and that these molecules are composed of atoms. In a few cases a molecule contains only one atom, or the molecule and the atom are identical. Broadly speaking, the science of chemistry is concerned with the study of the various elementary atoms and of their possible combinations. The science of physics deals with the properties of bodies in so far as they depend upon the peculiarities of their molecules.

In most physical operations the molecules of the bodies which are studied are unchanged by those operations. In chemical operations the molecules break up, and new molecules of other sorts are formed. The elementary atoms are not appreciably changed in any ordinary chemical operations, and, until very recently, they have always been assumed to be indestructible. But the study of radioactive bodies has shown it to be highly probable that, at least in certain cases, the atom is really a composite body, and that it may change its character by the loss of some of its parts. We may, for the present, assume the atom to be indestructible.

185. The Kinetic Theory of Gases. — The properties possessed in common by all gases indicate a common cause to which those properties may be ascribed. As far back as the year 1738 an attempt was made by Daniel Bernoulli to explain Boyle's law as the result of the purely mechanical impact of the particles or molecules of the gas. Bernoulli assumed that the molecules of a gas are in constant motion, and that they collide with one another and with the walls of the vessel which contains them. He also assumed that their number in any ordinary volume is enormously great. It is

evident that the effect of the collisions of such molecules with one another would be to alter the velocities of the individual molecules, so that, even if they were equal at any time, they would not long remain the same. For the purposes of elementary calculation it must be assumed that the effect which they would produce by their impacts against the walls of the vessel which contains them would be the same as that which they would produce if they all had a common velocity. The hypothesis asserts that the pressure of the gas on the walls of the containing vessel is due to the impacts of the molecules. The problem then is to determine, on these suppositions, how the pressure on the walls will depend upon the volume of the vessel and the velocity of the molecules

The following solution of it is the one given by Bernoulli: The gas is supposed to be contained in a cylindrical vessel (Fig. 128), and confined within it by a piston, which at first stands at unit distance from the base of the cylinder. If the piston is then pushed down until its distance from the base is s, the volume occupied by the gas may be represented by s, if its original volume is represented by 1. It is assumed that the common velocity of the molecules is not changed by this operation. Now there are two reasons why the pressure of the gas should be greater, when its volume is thus reduced, than it was before: first, the molecules have been crowded together into a smaller volume, and consequently there will be more molecules lying near the walls; secondly, the average volume occupied by any one molecule is diminished, and consequently the molecule will strike the side of this volume more frequently.

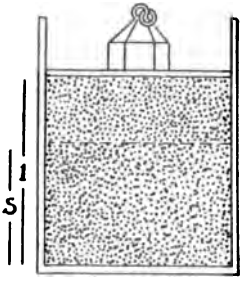


Fig. 128.

If we consider the first cause assigned, it appears that the number of molecules which will lie contiguous to unit area of the wall will be inversely proportional to the square of the cube roots of the volumes of the gas. This is seen if we notice that the number of molecules in unit volume will be inversely as the volumes of the gas in its two conditions, so that the number that will lie along unit length will be inversely as the cube roots of the volumes, and therefore the number that will lie on unit area will be inversely as the square of the cube roots of the volumes. Thus the ratio of the pressures in the two conditions of the gas, so far as they depend on the first cause, will be given by $s^{\frac{2}{3}}: 1$.

If we consider the second cause assigned, it appears that the number of impacts of any one molecule against the walls will be inversely as the average distance between the molecules, or inversely as the cube roots of the volumes of the gas. This is seen if we notice that the average distance between the molecules is inversely as the number that will lie along unit line, and that this number, as already stated, will be inversely as the cube roots of the volumes. Thus the ratio of the pressures in the two conditions of the gas, so far as it depends on the second cause, will be given by $s^{\frac{1}{3}}: 1$.

When we consider that both the causes assigned act together, we conclude

that the ratio of the pressures in the two conditions of the gas is given by $(s^{\frac{3}{2}})(s^{\frac{1}{2}}) : 1$, or by $s : 1$, and hence conclude as the result of our original hypothesis as to the nature of a gas, and the way in which it exerts pressure, that when the velocity of the molecules of a gas is maintained constant, its pressure varies inversely with its volume.

To show the way in which the pressure depends upon the velocity of the molecules, when the volume is unchanged, we notice that the impulse applied by each molecule to the wall, measured by its change in velocity, is proportional to its velocity, and that the number of impulses which each molecule exerts upon the wall in unit time is also proportional to its velocity, so that the effect due to any one molecule, and therefore the effect due to their combined action, is proportional to the square of their velocity. If we assume that the velocity of the molecules is increased by the introduction of heat, we may explain in this way the increase of the pressure which a gas exerts when its temperature rises, and by the use of Gay-Lussac's law we may obtain a simple relation between the velocity of the molecules and the temperature of the gas.

186. The Fundamental Equation of the Kinetic Theory. — The success of the kinetic theory in explaining the behavior of gases has been so complete, and its influence in establishing the general kinetic theory of matter has been so great, that it is worth while to discuss this question more formally, so as to obtain the fundamental equation of the kinetic theory of gases.

We assume that a gas consists of an assemblage of minute molecules, which are in constant motion, colliding frequently with one another and with the walls of the vessel which contains them. Although it is evident that the effect of such collisions would be to bring about different velocities in the different molecules, we shall assume, to facilitate the calculation, that they all have the same velocity. We suppose them perfectly elastic, so that no kinetic energy is lost by the collisions, and in particular, so that a molecule which strikes the bounding wall moves away from it with the same normal velocity as that with which it approached it. We also suppose that there are no forces exerted by one molecule on another, except at the instant of collision, and that the time of collision is negligibly small in comparison with the time between collisions. In the actual state of a gas, no one molecule will occupy the same part of the containing vessel for any length of time, but if it moves out of an element of volume its place will be taken by another molecule, and in this calculation we may assume that the whole volume occupied by the gas is divided into small equal elements of volume, as many as there are molecules of the gas, and that each one of these is occupied by a molecule. Those molecules which lie next the wall beat against it as they move

about, and the pressure of the gas is measured by the average force which is exerted by each unit of area of the wall in turning back the molecules colliding with it. On the suppositions which have been made we shall calculate this average force. It is equal to the total change in momentum experienced by the molecules contiguous to unit area of the wall, in the small time t, divided by that time. To get the change in momentum we need (1) the change in momentum experienced by a molecule in a single collision, (2) the number of collisions of a single molecule with the wall in the time t, (3) the number of molecules lying contiguous to a unit area of the wall, (4) the proportion of these which may be thought of as making normal impact against the wall.

(1) A molecule of mass m moving directly against the wall with the velocity u, rebounds from it with the same velocity, so that its change of momentum is $2mu$.

(2) To get the number of collisions we need to know the size of the elementary volume assigned to a molecule. To represent this we suppose that the total number of molecules is n and that the volume which they occupy is v; then the volume occupied by each molecule will be, on the average, v/n. This may be considered a cube, and the edge of it will have the length $(v/n)^{\frac{1}{3}}$. The molecule moving perpendicularly to the wall with the velocity u will pass twice through this length between collisions, so that what we may call the average time for one collision is given by $2(v/n)^{\frac{1}{3}}/u$, and the number of collisions in the time t by $tun^{\frac{1}{3}}/2v^{\frac{1}{3}}$.

(3) The number of molecules which will lie contiguous to unit area is found from the number of the elementary cubes, each containing a molecule, which can be set on unit area. The area of a face of one of these cubes is $(v/n)^{\frac{2}{3}}$, so that the number of them which will stand on unit area is $(n/v)^{\frac{2}{3}}$

(4) If we multiply together the results of the preceding paragraphs we obtain the total change of momentum experienced during the time t by all the molecules contiguous to unit area of the wall, on the supposition that they are all moving perpendicularly against it with the velocity u. On so doing we obtain mnu^2t/v. This cannot be the correct solution of the problem, for the supposition that all the molecules are moving perpendicularly to the wall is evidently erroneous. The molecules are moving in every direction, and the velocity with which they approach the wall is the average component of the velocity u in the direction perpendicular to the wall. If we write c_1, c_2, c_3 for the three rec-

tangular components of the velocity u of any one molecule, of which c_1 is the component perpendicular to the wall, we have $u^2 = c_1^2 + c_2^2 + c_3^2$, and the average values of these components for the whole number of molecules will conform to the same equation, in which now the symbols represent average values. But when a sufficiently large number of molecules is considered, the probability is that their average velocity in any one direction will be the same as in any other, so that we may set $c_1^2 = c_2^2 = c_3^2$; or $u^2 = 3c^2$. It therefore appears that we should multiply the change of momentum obtained above by the numerical factor $\frac{1}{3}$, if we retain u to represent the average velocity, in order to represent the true value of the change of momentum experienced by molecules moving in every direction.

The force on unit area measures the pressure. Representing it at once by p, the impulse applied by the wall to the molecules colliding with it in the time t is pt, and this is equal to the change of momentum experienced by the molecules, so that we have finally

$$p = \frac{1}{3}\frac{mnu^2}{v}. \tag{80}$$

This equation is the fundamental equation of the kinetic theory of gases.

Boyle's law follows immediately from this equation. For if we write it $pv = \frac{1}{3}mnu^2$, the quantities on the right are all constant, so long as conditions are maintained so that the average velocity of the molecules remains the same. The theory assumes that this velocity changes with the temperature of the gas, and remains the same so long as the temperature is constant. On that condition, therefore, which is the necessary condition for the applicability of Boyle's law, the kinetic theory indicates that the product of the pressure and the volume of any mass of gas should be constant.

We may extend this demonstration somewhat by making a hypothesis about the velocities of the molecules, which more nearly represents their true velocities than the one which we have hitherto used. We suppose that the velocities are not the same, but different. The number of molecules is, however, so great that there will still be a large number, n_1, whose velocities lie very near a common value u_1. This set of molecules will exert a partial pressure p_1 given by the formula $p_1 v = \frac{1}{3}n_1 m u_1^2$. Similar partial pressures will be exerted by other sets of molecules, whose velocities lie near the values u_2, u_3, u_4, etc. The sum of all these partial pressures is the pressure of the gas. If we represent by n the number of molecules of the gas, by u a certain average velocity, by n_1, n_2, n_3, etc., the numbers of molecules in the different sets which have been considered, we may write an equation defining the average velocity as follows: $nu^2 = n_1 u_1^2 + n_2 u_2^2 +$

$n_1u_1^2 +$, etc. The average velocity thus defined is called the velocity of mean square. In terms of this velocity we obtain the equation $pv = \frac{1}{3}nmu^2$ for the pressure of the gas.

As has already been remarked, the equation just obtained expresses Boyle's law. As will be seen later, when taken in connection with Gay-Lussac's law, it affords us a measure of temperature and an insight into the nature of heat. By the aid of an additional theorem, first given by Maxwell, it may be made to demonstrate Avogadro's law also. Maxwell proved that, when two gases are mixed, their most probable condition, or their condition of final stability, is that in which the mean kinetic energies of the molecules of both gases are equal. This being so, let us consider equal volumes of two gases under the same pressure and at the same temperature, so that they conform to the condition of being in mechanical and thermal equilibrium with each other. In these circumstances the product $pv = \frac{1}{3}nmu^2$ is the same for each. But by Maxwell's law the factor mu^2 is also the same for each, so that the factor n will be the same for each. Hence Avogadro's law follows as a consequence of the kinetic theory.

It was shown by Joule that we may calculate the velocity of mean square of the molecules of a gas from the fundamental equation. If we choose $v = 1$ in that equation, the product mn represents the mass in unit volume or the density of the gas. Using ρ to represent the density of the gas, the equation becomes $p = \frac{1}{3}\rho u^2$. The pressure p corresponding to this density can be measured and hence the value of the velocity calculated.

Since the velocities for all gases can be calculated from the velocity for hydrogen, it is worth while to calculate this. The density of hydrogen under the pressure of one atmosphere and at 0° C. is $\rho = 8.954 \times 10^{-5}$ grams per cubic centimeter. The pressure of one atmosphere equals 1.013×10^6 dynes per square centimeter. With these data we find u for hydrogen equal to 1842 meters per second. This is a little more than a mile per second. From Maxwell's law of the equality of the kinetic energies, the velocity of the molecule of any other gas is to the velocity of the hydrogen molecule inversely as the square root of the mass of the molecule of the gas is to the square root of the mass of the hydrogen molecule. Thus if the mass of the oxygen molecule is sixteen times that of the hydrogen molecule, the velocity of the oxygen molecule is one-quarter that of the hydrogen molecule.

187. The Radiometer. — The free motion of the molecules of a gas serves to explain the operation of an instrument invented by Crookes, and named by him the radiometer. In one of its forms the *rodiometer* consists of a thin disk, or vane of mica, carried on

the end of a light horizontal rod, which is suspended by a long fine wire or quartz fiber. The vane is so placed that pressure on one of its faces will cause a twist in the wire. One of the faces is blackened with lampblack, so that it will absorb heat radiation freely, while the other is left polished. This apparatus is enclosed in a glass chamber, from which the air is exhausted about as well as can be done with a good air pump. When it is placed so that light, or radiant heat, falls on the blackened face of the vane, the system moves as if a pressure had been applied to that face.

To explain this motion we consider that in the rarefied state of the air in the chamber the free paths of the molecules are very long in comparison with those in ordinary air, and the molecules which strike against a face of the vane often pass to the wall of the chamber without encountering others. There is therefore no such opportunity as there is in air at ordinary pressures for the velocities of the molecules to be frequently exchanged and altered, so as to produce a uniformity of conditions in all parts of the chamber. When the blackened face of the vane is heated by the absorption of radiant heat, it communicates greater average energy to the molecules which strike against it than is communicated by the polished face to those which strike against it. The molecules are therefore flung off from the blackened face with a greater average momentum than from the polished face, and their reaction against the blackened face is greater, so that it moves as if a pressure were applied to it.

At ordinary densities the increased energy of the molecules flung off from the heated surface is so quickly distributed by collisions throughout the whole mass of air in the chamber, that the velocities of the molecules are the same on both sides of the vane, and there is no difference of pressure between the two faces.

188. The Brownian Movements. — Gases do not seem to differ essentially from liquids or from solids, except in being far less dense, so that the average distance apart of their molecules is greater. If the molecules in a gas are in motion, it seems probable that those in a liquid or solid are in motion also. The evidence in favor of this view will be considered in connection with our study of the effects of heat on bodies. Owing to the difficulties of calculation introduced by the fact that in these bodies the molecules are always within one another's range of action, it has not been found possible to develop a kinetic theory of liquids, and much less one of solids. One general result has, however, been reached by the

labors of Maxwell and of Boltzmann, which is known by the name of the Maxwell-Boltzmann theorem.

This theorem states that in any assemblage of particles in which the temperature is everywhere the same, the average kinetic energy associated with each degree of freedom of each particle is the same. If applied only to the kinetic energy of translation of each particle, it states that the average kinetic energy of each freely moving particle in the assemblage is the same. A special form of this theorem, proved by Maxwell, was used in § 186 to demonstrate Avogadro's law.

A confirmation of this theorem for liquids has recently been obtained by the study of the so-called *Brownian movements*. If finely divided matter, like carmine, India ink, or gamboge, is mixed with water, so that its particles are in suspension, an examination of them with a powerful microscope will show that they are all executing irregular motions. These motions were first discovered by Dr. Brown in the small particles in certain vegetable cells, and were taken by him as evidence of the life of the cell, but the discovery that they are equally well exhibited by inorganic particles made it necessary to find a mechanical explanation of them. They are not due generally to the impact of any one molecule of water against the particle, for a calculation of the momentum of a molecule of water shows that it would be too small to account for the movements observed. We may give an explanation by taking into account the fact that while the average energy of any one molecule is in the long run the same as that of another, and while the average energy of any one fairly large group of molecules is generally the same as that of another, yet the actual velocities of the molecules are not the same at all times or in all places. If we fix our attention on a very small group of molecules, we shall find that its energy, and therefore the velocity of its constituent molecules, will change from time to time, and will not be the same at any one time as the energy of another similar group.

Now consider a small solid particle surrounded by water. The molecules of water which are near it may be divided into numerous small groups, and the energies of these groups will not be always the same. The molecules which strike it all over its surface will have different momenta, of which the resultant will not always vanish. It may be great enough to give rise to a force sufficient to move the particle. The particle will therefore be impelled in one direction or another from time to time by this resultant mo-

mentum. The probability of the occurrence of such inequalities in the momenta imparted to the particle by the surrounding molecules was worked out by Smoluchowski, and found to be competent to account for the actual motions observed.

The distribution of the molecules of a gas in a column acted on by the force of gravity can be calculated on the principles of the kinetic theory, and a similar calculation employed to determine the distribution of much larger bodies. Perrin has experimentally verified that the calculated law of distribution holds in the case of particles of gamboge in water, and has further shown that the energy of translation of the particles is the same as that of a molecule of gas at the same temperature, and has thus confirmed the Maxwell-Boltzmann theorem.

It should be said that the verification of this theorem encounters great if not insuperable difficulties when we attempt to apply it to the parts of the atoms of matter which are concerned with the emission or absorption of heat and light.

CHAPTER IX

FRICTION

189. Resistance to Motion Due to Friction. — When a block of iron or wood is pushed over a table, the moving force is always opposed by a resistance, or force in the direction opposite to the motion. This resisting force may be called the force of friction, or simply the *friction* between the two bodies which slide over each other. The magnitude of the friction depends upon the nature of the bodies and upon the state of their surfaces, smooth bodies experiencing less friction than rough ones. A solid of any form moved through a liquid, or through a gas, encounters a similar resistance. In this case the parts of the fluid which lie nearest to the solid are set in motion by it, and as they in turn slide past other parts of the fluid, they experience friction also, and communicate their motion to those parts, so that the work which is done in moving the solid against friction is largely spent in moving the parts of the fluid. When the motion of the solid ceases, the moving fluid is gradually brought to rest by the friction between its parts. The friction between the parts of the fluid is called *internal friction* or *viscosity*. The fact that bodies moving through the air are resisted by the friction of the air was known to Galileo, who used it to explain the fact that very light bodies with large surfaces do not fall at the same rate as heavy compact bodies do.

190. Friction Between Solids. — The frictional resistance which a solid body encounters when it is slid over a solid surface is believed to be due partly to the direct action of forces between the molecules of the two bodies, and partly to the partial interlocking of the small protuberances on the two surfaces. The latter cause is more prominent when the surfaces of the bodies are rough; the former is, at least relatively, more prominent when their surfaces are smooth or polished.

As has already been explained in § 46, a solid at rest on a solid surface offers a resistance to a force applied to it parallel to the surface, and does not start moving until the applied force exceeds a certain value. This static friction, which is exhibited only by

solids, is explained in the same way as the friction exhibited by one solid sliding over another.

191. Laws of Sliding Friction. — The laws of sliding friction were investigated by Coulomb. One of the two bodies which he employed was in the form of a long horizontal plane, like a flat board. The other body was pulled over it by a weight, and the observation consisted in determining the value of the weight required to give the moving body a constant velocity in different circumstances. He found that the friction between two bodies is proportional to the force pressing them together, that it is independent of the extent of surface in contact, and that it is independent of the velocity of the moving bodies.

192. Laws of Rolling Friction. — Friction is also exerted upon a body which rolls over another one, so that, for example, friction is exerted upon a wheel rolling along a surface. Coulomb found, for the case of rolling friction, that the friction is proportional to the force pressing the bodies together, inversely proportional to the radius of the wheel, and independent of its velocity.

193. Coefficients of Friction. — In the case of sliding friction, the friction or the resistance is measured by the force required to maintain a constant velocity in the moving body. The ratio of this resistance to the normal force on the body is called the coefficient of sliding friction. The coefficient of rolling friction is given by an essentially similar definition.

These laws of friction are only approximately true. It has been found that the friction falls off very considerably when the velocity of the moving body is very great. Its value probably changes also when the pressure is great. The coefficient of friction is very much diminished by the use of lubricants.

194. Friction of Liquids. — The internal friction of a liquid was discussed theoretically by Newton, on the assumption that it depends on the relative velocities with which contiguous sheets of the liquid slide past each other. This sort of motion is that which is produced by a shearing stress. Since all liquids are viscous, that is, since they all offer more or less resistance to this sort of motion, it cannot be said that liquids offer no resistance to a shearing stress. The resistance which they do offer, however, depends on the relative motion of their parts, and since liquids do not possess true rigidity, any shearing stress acting in them will produce motion, and the liquid will gradually assume a form in which it is free from shearing stress.

By the experimental and theoretical study of the effect of friction on the motion of a solid in a liquid, or of the motion of a liquid flowing past a solid wall, or through a tube, it has been found that the motion is affected as if the portions of the liquid nearest the solid adhered to it without slipping, so that the resistance to the motion is due entirely to the viscosity of the liquid. When the flow of a liquid through long capillary tubes is treated on this assumption, it may be shown that the amount of liquid which issues from the tube in the unit of time is proportional to the fourth power of the radius of the tube. This theoretical conclusion is in accord with the results of experiments carried out by Poisseuille.

The resistance to the motion of a solid through a liquid increases with its velocity. When a force of a given value is applied to a body to move it through a liquid, the velocity of the body will gradually increase until it reaches a certain limiting value, for which the resistance due to friction is equal to the applied force. After that value has been reached, the velocity will remain constant so long as the force is applied. Thus a body falling through water will have a constant velocity after it has fallen a certain distance.

195. Friction of Gases. — In general what has been said about the internal friction of liquids applies also to gases. The peculiarity of the case of gases is this, that we can explain their viscosity by means of the kinetic theory of gases. When one layer or sheet of a gas slides past another, the molecules which dart out from the more rapidly moving sheet into the other one, carry with them a certain momentum in the direction in which that sheet is moving, and thus communicate momentum to the more slowly moving sheet. The molecules which dart out from the more slowly moving sheet into the other one communicate to it a certain momentum in the direction opposite to its motion. Thus the momenta of the two sheets tend to become the same, or the sheets exert a force on each other. Starting with this conception, the laws of internal friction in gases can be deduced, and shown to be consistent with those obtained by experiment. In particular, Maxwell, to whom the development of this theory is due, showed that the internal friction of gases should be independent of their densities. The experiments which Maxwell and others carried out to test this conclusion showed that it holds true within very wide limits. It fails only when the density becomes exceedingly small.

The resistance offered by a gas to the motion of a body through it depends on the velocity of the body. When the velocity is small, like that of a swinging pendulum or a magnet, the results of experiment can be represented by supposing the friction to be proportional to the velocity. For more rapidly moving bodies, the friction seems to be nearly proportional to the square of the velocity. For bodies moving with the high velocities of the modern rifle ball or cannon ball, the friction seems to be even greater than this law would indicate.

The theory of the motion of a sphere through a gas leads to the conclusion that the force required to overcome the retarding effect of friction, and to keep the sphere moving with a constant velocity, is proportional to the radius of the sphere. If the sphere is falling through the gas, the force which moves it, or its weight, is proportional to the cube of the radius. It will fall with continually increasing velocity until it reaches at last a limiting velocity, for which its weight is equal to the resistance offered by friction. This limiting velocity is therefore proportional to the square of its radius. Thus the limiting or constant velocity attained by large spheres falling through the air will be greater than that attained by small spheres. Large drops of water which fall from a cloud will have considerable velocity when they reach the ground, while small drops will move toward the ground more slowly. The very small drops which form a fog or cloud will fall so slowly that their motion is hardly perceptible, and can easily be reversed or altered by the motion of currents of air.

CHAPTER X

DIFFUSION

196. Solution. — If a lump of sugar is dropped into a vessel of water, it gradually disappears. The process by which it disappears seems to be a gradual disintegration of those parts of it which are nearest the water, and an absorption of them into it. This process is called *solution*, and the sugar is said to dissolve in the water. After solution the water has acquired certain properties which it did not possess before. These are supposed to be due to the presence of molecules of sugar distributed throughout the water.

Very many bodies will dissolve in water or in other liquids. The liquid in which solution takes place is called the *solvent;* the body which dissolves in it, the *solute.* A limited amount of solvent will not dissolve an unlimited amount of solute. After a certain amount has been dissolved, the limit of solution has been reached, and if more of the solute is present, it will remain undissolved. The amount of solute which can be dissolved in unit mass of the solvent is called its *solubility.*

In some cases there seems to be no limit to the amount of solute which will be taken up by a solvent. Thus a homogeneous mixture may be formed by adding alcohol or glycerine in any quantity to a limited quantity of water. In cases of this sort we speak of the one body as dissolved in the other when only a relatively small amount of it is used, but in general the two bodies are said to be mixed, and the body formed is called a mixture.

We may explain the act of solution, in a general way, by supposing that the molecular forces, acting between a molecule of the solute and the neighboring molecules of the solvent, are strong enough to tear away the molecule of the solute from its original position and to transfer it to the solvent. It is uncertain whether the molecule of the solute, after it has entered the solvent, is free, or whether it is permanently bound to one or more molecules of the solvent, so as to form what is called a molecular aggregate. The weight of evidence is at present in favor of the existence of such molecular aggregates in many, if not in all, solutions.

197. Free Diffusion of Liquids. — If a quantity of sulphuric acid is placed in the bottom of a tall cylindrical jar, and if water is then poured carefully into the jar in such a way that the sulphuric acid below is not disturbed, the water, being relatively lighter, will float upon the acid, and the surface of separation between the two liquids will be distinct. In process of time this distinct surface disappears, and a region exists between the pure acid below and the pure water above in which there is a mixture of water and acid. This region gradually extends until at last it occupies the whole space filled by the liquids. At first the proportion of the two liquids found in different parts of the column is very different, but as time goes on the mixture becomes more and more homogeneous. In theory, however, it cannot become entirely homogeneous until after an infinite time has elapsed. This process, by which one liquid mixes with another, while the mixture is not aided by currents set up in the mass, is called *diffusion*, and a case of this particular sort, in which the two liquids are not separated by any third body through which they must pass to mix, illustrates free diffusion.

An experiment which is essentially similar may be tried by placing pure water upon a solution of copper sulphate. The dissolved salt will gradually rise through the pure water, until the whole mass becomes a solution of copper sulphate of uniform strength.

Any two liquids which may be mixed will diffuse into each other in the way here described. When one of the two is a solution, we consider the process to be the diffusion of the solute from the parts of the mass in which its concentration is greater to the parts in which its concentration is less. It was assumed by Fick, and his assumption has been verified as approximately correct in many cases, that the rate at which diffusion takes place along a line is proportional to the rate of change of concentration along that line.

We may explain diffusion by ascribing it to a force exerted upon the molecule of the solute by the molecules of the surrounding solvent, it being supposed that those portions of the solvent in which there are fewer molecules of the solute exert the greater force. The motion which this force will set up is resisted by the friction experienced by the molecule of the solute as it passes through the solvent. Certain experiments which have been made indicate that this friction is very great.

The process of diffusion may be explained more satisfactorily by supposing it to be due to the motions of the molecules of the solute.

We have every reason to believe that these molecules are constantly in motion, and that their motion is in general similar to that already considered in our study of gases. It is easy to see that the effect of such motions would be to distribute the molecules of the solute throughout the whole volume to which it has access.

198. Diffusion through Membranes. — The earliest systematic study of diffusion was made on the diffusion of solutions through solid substances. Many substances, as, for example, unglazed earthenware, and organic bodies, like parchment paper or a bladder, or even india rubber in thin sheets, will permit liquids to diffuse through them. If a solution of common salt and water is placed in a bag made of parchment and immersed in a vessel of water, the salt will gradually pass out through the bag, and water will enter it. If another salt is dissolved in the water outside the bag, it will generally enter the bag during the process. The rates at which the two salts pass through the membrane depend upon the nature of the salts and of the membrane. The general process here described was called by Dutrochet *osmosis*.

In studying the phenomena of osmosis, Graham found that bodies differed very remarkably with respect to the facility with which they can pass through an organic membrane. He divided all soluble bodies into two classes, to which he gave the names crystalloid and colloid. *Crystalloid* bodies, like sugar or salt, are those which in their solid state have generally a recognizable crystalline form. When dissolved in water, they diffuse through an organic membrane almost as freely as if the membrane were not present. *Colloid* bodies, like gum arabic or glue, are amorphous, or have no crystalline form. They scarcely diffuse at all through an organic membrane. This distinction between crystalloids and colloids also appears in the case of free diffusion, the colloids diffusing very much slower than the crystalloids. Graham based on this difference a process, called *dialysis*, of separating salts from the colloid organic bodies with which they may be mixed.

199. Laws of Osmotic Pressure. — In general, as has already been described, osmosis takes place in both directions through an organic membrane, and with such membranes no very exact or general laws of osmosis can be discovered. Traube discovered that membranes can be obtained, by taking advantage of chemical action, which in many instances permit osmosis in only one direction. Such membranes are called semi-permeable membranes. If a small mass of sulphate of copper is dropped into a dilute solution

of ferrocyanide of potassium, a film of cyanide of copper forms over its exterior. Water can pass by osmosis through this film, but most substances which dissolve in water do not pass through it. A film of this sort may be deposited inside the pores of an unglazed earthenware jar. If the jar is filled with a solution of sugar, to take a specific example, and is partially immersed in water, the sugar will remain in the jar and water will enter the jar from without. The membrane is permeable to water, but is not permeable to sugar.

The first experimental study of osmosis through semi-permeable membranes was made by the botanist Pfeffer. He used a small jar of the sort just described, closed tightly at the mouth by a stopper, through which passed a long glass tube. This arrangement we call an osmotic cell. When such a cell was filled to the stopper with a solution of sugar, or of any ordinary salt, and was immersed in water, the water gradually entered the cell, so that a column of liquid rose in the tube. This process continued for many days, but it ceased at last when the column attained a height depending upon the concentration of the solution. In any such experiment the height of the column thus attained measures the pressure which will force water out through the membrane as fast as it is coming in by osmosis. The pressure thus measured is called the *osmotic pressure* of the solution. If we arrange an osmotic cell in such a way that pressure can be applied to the surface of the solution in it, and so adjust this pressure, when the cell is immersed in the water, that water neither enters nor leaves the cell, the pressure thus determined is the osmotic pressure.

By the study of the results obtained by Pfeffer, van't Hoff has shown that the magnitude of the osmotic pressure depends upon the amount of solute in the solution, in the same way exactly as the pressure of a gas depends upon the amount of gas enclosed in a given volume.

In the first place, the osmotic pressure of a solution of a given solute is proportional to the concentration of the solution; that is, to the amount of solute in unit volume of the solution. If we consider the same quantity of solute in solutions of different strengths, it is plain that the concentrations of those solutions are inversely as their volumes. Thus the law just stated is equivalent to the statement that when the quantity of solute is fixed, the osmotic pressures of solutions of different strengths are inversely as their volumes. That is, the osmotic pressure obeys a law like Boyle's law.

In the second place, for a solution of given strength, the osmotic pressure rises with the temperature according to the same law as that which determines the relation between the pressure and temperature of a gas. That is, the osmotic pressure obeys a law like Gay-Lussac's law.

In the third place, the osmotic pressures of many solutions are equal when their molecular concentrations, measured by the number of molecules of the solute contained in unit volume of the solution, are equal. That is, for these solutions, the osmotic pressure obeys a law like Avogadro's law. In many other cases, however, this law is not fulfilled. Arrhenius explained this fact by supposing that some of the molecules of the solute, in these cases, are disintegrated into their constituent parts. The number of independent portions of the solute — they cannot now be called molecules — is thereby increased, and the osmotic pressure increases likewise. The molecules thus broken up are said to be *dissociated* or ionized. This hypothesis not only accounts for the osmotic pressures exhibited by such solutions, but also is consistent with and explains many other of their peculiarities.

200. Diffusion of Gases. — The free diffusion of gases goes on in a way essentially similar to that of liquids. Its fundamental laws have been developed by Maxwell from the kinetic theory.

The diffusion of gases through porous plates of earthenware, gypsum, or graphite, was studied by Graham, who showed that the rate of diffusion is proportional to the difference of the pressures on the two sides of the porous plate, and is inversely proportional to the square roots of the relative densities of the gases. Thus oxygen, which is sixteen times as heavy as hydrogen, will diffuse only one-quarter as fast. By taking advantage of this fact, we may partially separate mixtures of different gases. These laws have been shown by Reynolds to be consistent with the conclusions of the kinetic theory.

Gases will diffuse through thin films of liquid, like the walls of a soap bubble. They will diffuse also through certain metals, when they are heated. Thus coal gas will diffuse through the fire pot of a furnace, if it becomes red hot.

BOOK III

SOUND

CHAPTER XI

SOUND AND SOUND WAVES

201. General Considerations Respecting Sound. — The sensation of sound, given to us by our sense of hearing, is perfectly familiar to every one. While the immediate sensation has no apparent connection with matter, yet universal experience has shown that sounds are transmitted through or by matter from bodies in which they originate. The science of sound is concerned with the study of the actions in sounding bodies and with the mode of transmission of these actions through space.

From the time of Aristotle it has been commonly believed that sound is transmitted through the air as a series of disturbances or impulses communicated to it by the body which is the source of the sound. A rival theory, which assumed that sound is transmitted by minute particles emitted from the sounding body, had so little to support it that it was never given serious consideration. All subsequent study has shown that Aristotle's theory of the origin and the mode of transmission of sound is correct, and the progress of knowledge has resulted simply in more exact statements regarding the mode in which the sounding bodies and the transmitting medium act.

Confusion sometimes arises because the word sound is used to designate a sensation received through our sense of hearing, as well as the physical action by which that sensation is excited. In what follows we shall commonly use the word to denote the physical action, which can often be detected even when it produces no audible sound.

202. Musical Sounds. — Certain sounds produce in us a peculiar sensation which so distinguishes them from others that they are called *musical sounds* or *tones*. Such musical sounds excite a sensation which lasts a perceptible time without perceptible change.

In many cases also the sensation produced possesses a certain simplicity which distinguishes it from sounds of other sorts. Other sounds, which do not possess these characteristics, are called noises. We shall confine our attention almost altogether to musical tones.

203. Sounds Produced by Strings. — The use of stringed musical instruments is of great antiquity. It must have been observed, almost as soon as such strings were used, that the tone produced by the string depends in some way on the force or tension with which the string is stretched and on the length of the string. It is credibly reported that Pythagoras investigated the relation between the length of a string and the pitch of the tone emitted by it, and found that when strings whose lengths were to one another as one, one-half, two-thirds, and three-fourths, were under the same tension, the tones emitted by them were those of the ordinary musical scale called the fundamental, the octave, the fifth, and the fourth. This discovery of Pythagoras is the first one recorded in the history of physics, antedating by more than two hundred years any recorded observations in mechanics.

Mersenne (1588–1648) extended this observation of Pythagoras by determining the lengths of string which produce all the tones of the ordinary musical scale. He also recognized that the time taken by the string to execute one vibration, called its *period*, or, what amounts to the same thing, the *number of vibrations* executed by the string in one second, called the *frequency*, which is the reciprocal of the period, depends on the length of the string. He supposed that the period is directly proportional to the length of the string, or that the frequency is inversely proportional to the length of the string. We may therefore state the relation which Mersenne determined between the tones of the musical scale and the lengths of the strings which produce them as a relation between those tones and the frequency producing them. These relations are given for the major and minor scales in the following tables:

MAJOR SCALE

Name	do	re	mi	fa	sol	la	si	do
Number of vibrations	n	$\frac{9n}{8}$	$\frac{5n}{4}$	$\frac{4n}{3}$	$\frac{3n}{2}$	$\frac{5n}{3}$	$\frac{15n}{8}$	$2n$

MINOR SCALE

Name	la	si	do	re	mi	fa	sol	la
Number of vibrations	n	$\frac{9n}{8}$	$\frac{6n}{5}$	$\frac{4n}{3}$	$\frac{3n}{2}$	$\frac{8n}{5}$	$\frac{9n}{5}$	$2n$

On the assumption that the length of a string is inversely proportional to its frequency, Mersenne determined the frequency of a string emitting a particular tone, by comparing its length with that of a string so long that its vibrations could be counted.

In all that has been said so far, the supposition has been made that the strings are under the same tension. The pitch of the tone emitted by a string depends, however, not only on its length, but on its tension, on the material of which it is composed, and on its thickness. Mersenne found, for strings of the same material, that the frequency is proportional to the square root of the tension, is inversely proportional to the length, and is inversely proportional to the thickness. Later study has shown that if strings of different materials are used, the frequency is proportional to the square root of the tension, is inversely proportional to the length, and is inversely proportional to the square root of the mass of the string in unit length.

204. Transmission of Sound. — After it had thus been shown that a musical tone originates at a body which is executing regular vibrations, attention was turned to the question of the transmission of these vibrations. They are manifestly transmitted, in most cases, through the air. Other bodies however, may serve as the medium of transmission.

Thus the vibrations of a piano string may be transmitted to the auditory nerve through a wooden rod, one end of which rests on the lid of the piano, while the other end is held in the teeth. The vibrations pass through the rod and through the bones of the head, without passing through air. A scratch of a pin, or a light blow of a hammer on one end of a long log, is heard twice by a person at the other end. One of the sounds is transmitted through the wood, the other is transmitted less rapidly through the air.

The problem of the mode of transmission of sound was first successfully studied by Newton. He assumed that the sounding body emitting a musical tone is vibrating in such a way that the motion of each of its parts is a simple harmonic motion (§ 63). He assumed further, that vibrations of a similar sort are transmitted to the particles of air which are nearest to the body, from these to the next set of particles, and so on. He showed that, on these assumptions, the motion of any one particle of air will be a simple harmonic motion of the same period as that of the sounding body, and that the motion transmitted from particle to particle will proceed outward with a definite velocity.

We may describe the motion in the air by which sound is transmitted, in the following way (Fig. 129). Consider a long cylinder or tube filled with air, and suppose that a layer of air at one end of it is pushed forward. As it advances toward the next layer of air, it exerts a pressure upon it, which sets that layer in motion. The second layer in turn sets the third layer in motion, the third layer the fourth, and so on; and thus a forward motion is transmitted along the column. If the successive layers of air were rigidly bound together, the forward motion of the first layer would be transmitted instantaneously through the whole column. In fact, the transmission of the disturbance is not instantaneous, because the air is not incompressible. The first layer in the column moves up toward the second, the second up toward the third, and so on, producing a region in which the air is more condensed than it is in its condition of equilibrium. If

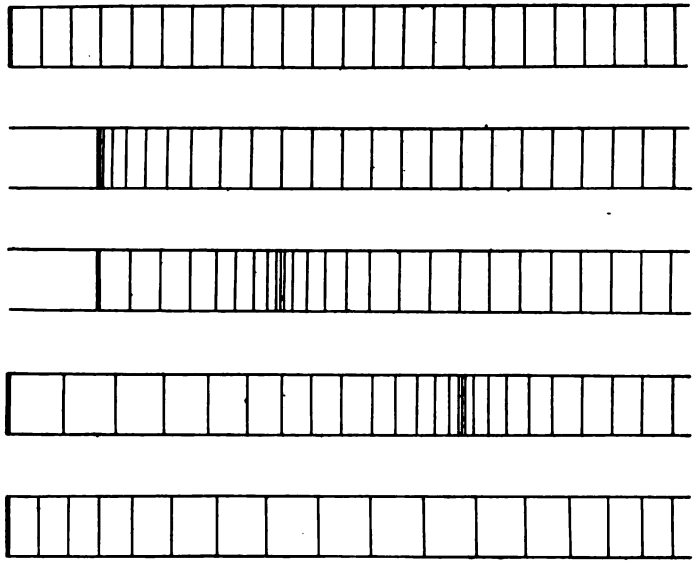


Fig. 129.

the forward motion of the first layer is checked after it has moved through a small distance, this process of condensation ceases, and in turn the other layers take up positions in which they are as far removed from one another as they were before the disturbing motion took place. The condensation originally produced proceeds as a form of motion along the column.

If the first layer is now suddenly drawn backward to its first position, the second layer is pushed toward it by the pressure on the other side of it, the third layer is pushed toward the second, and so on. The movements of these layers likewise take place successively, so that while they are going on, the air between the first layer and a part of the column which the disturbance has not yet reached occupies a larger volume than it does in its undisturbed condition, or is rarefied. When the first layer has come to rest, the other layers successively come to rest in the positions which they originally occupied, and the rarefaction which has been produced proceeds along the column.

Now, if the first layer of air is forced to execute regular vibrations, it will produce condensations and rarefactions similar to those which have been described, which will succeed each other at regular intervals and proceed along the column at a uniform rate as sound waves. The time in which the exciting vibration is executed, or the time between the production of two conditions of maximum condensation, is called the *period* of the sound. The distance between two successive points of maximum condensation is the *wave length.*

When a sound is produced, as is commonly the case, in a body of air extending in all directions, the successive condensations and rarefactions proceed outward from the body in all directions, and a particular condensation or rarefaction, which has left the body at any instant, will be found, at a later instant, on the surface of a sphere whose center is in the sounding body. The analogy between these spherical sound waves and the circular waves which are set up on the surface of still water, by a slight disturbance made at a point in it, was perceived by Vitruvius, although of course he had no exact knowledge of the character of sound waves.

205. Formula for a Simple Wave. — In case the waves in air originate at a source which is vibrating with a simple harmonic motion, the layers of air also vibrate with a simple harmonic motion of the same period, and any constant value of the displacement is transmitted through the air with a constant velocity. We shall proceed to represent by a formula, in the simplest case, the way in which the displacements of the layers of air affected by the passage of a series of waves depend upon the time and the distance from the source of the wave. The formula applies strictly only to waves traveling along a column of air, and is to be slightly modified when applied to spherical waves.

Let us represent by ξ the displacement of one of the displaced layers of air from its position when in equilibrium in the undisturbed condition of the air. The distance of this layer of air from the origin is represented by x, and the time that has elapsed since the displacement of the layer at the origin was a positive maximum is represented by t. Then if we write

$$\xi = a \cos 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right), \tag{81a}$$

in which a, T, λ are constants, we have a formula which represents a simple wave. For any constant value of x the motion of the layer

of air is a simple harmonic motion (§ 63), of the period T and with the amplitude a. The velocity v of the layer of air is given by

$$v = -\frac{2\pi}{T} a \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right), \tag{81b}$$

and the acceleration f, by

$$f = -\frac{4\pi^2}{T^2} a \cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right). \tag{81c}$$

For any definite value of t, the successive values of ξ along the axis of x vary with x as $\cos 2\pi(t/T - x/\lambda)$ varies. The maximum value of ξ is a, and the successive maxima are alternately positive and negative. If we compare the displacements at two points distant from each other by λ, we find that they are equal. The length λ is called the *wave length*, and the displacements at any instant repeat themselves at intervals of a wave length. If we use ordinates y perpendicular to the axis of x to represent numerically and in sign the displacements ξ, which are really to-and-fro displace-

Fig. 130.

ments along the axis of x, the curve (Fig. 130) represents the instantaneous condition of the air transmitting the waves.

If we determine a particular value of ξ at a certain point x_0 and at the time t_0, we see that this same value will be obtained at a later time $t_0 + \tau$, at a point distant from x_0 by such a distance χ that

$$\frac{t_0}{T} - \frac{x_0}{\lambda} = \frac{t_0 + \tau}{T} - \frac{x_0 + \chi}{\lambda}.$$

This gives $\frac{\chi}{\tau} = \frac{\lambda}{T}$ and χ/τ represents the velocity with which the constant value of ξ proceeds along the x axis. This velocity is the velocity of the wave. As the last formula shows, it is equal to the wave length divided by the period.

It must be understood that the formula which we have been discussing only serves to represent the characteristic displacements occurring during the passage of a succession of simple waves, and does not explain why such waves can be set up and maintained in air or in other elastic media.

Similarly the formula

$$\xi = a \cos 2\pi \left(\frac{T}{t} + \frac{x}{\lambda} \right) \tag{82}$$

represents a simple wave similar to the one we have been considering, except that it is traveling in the opposite direction. To show this let us use, as before, τ to represent a small time interval, and χ a distance measured along the x-axis. Then if a particular value of ξ is chosen, say for the time t_0, and at the point x_0, this value will be found, at the later time $t_0 + \tau$, at a point $x_0 + \chi$, given by the equation

$$\frac{t_0}{T} + \frac{x_0}{\lambda} = \frac{t_0 + \tau}{T} + \frac{x_0 + \chi}{\lambda}.$$

From this equation we get $\frac{\chi}{\tau} = -\frac{\lambda}{T}$, or the velocity of the constant value of ξ, or of the wave, is negative, or opposite to the velocity of the other wave form, and is of the same numerical value.

206. Dynamical Equation of Wave Transmission. — We shall next proceed to investigate the dynamical conditions in air or in any similar medium when a disturbance like that described in § 204 is set up in it. Let us consider the air at some particular instant, at which the different displacements of the layers of air along the x-axis are represented as in § 205 by the ordinates of a continuous curve (Fig. 131).

Fig. 131.

We first fix our attention upon two plane layers of air perpendicular to the x-axis, one of them originally at the position designated by x, the other at the position designated by $x + c$, where c is a small distance. In the disturbed state of the air the first layer is displaced by ξ, corresponding to x, the second layer is displaced farther by a very small distance that depends on c. If we use e to designate the rate at which ξ changes as x increases, we have ec for the amount by which the second layer moves farther than the first one.

We next fix our attention on two similar plane layers, a little farther along the axis, at the small distance b from the layers just considered. In the disturbed state of the air the second layer will move farther than the first one, by an amount that depends on c, but the amount by which the second layer moves farther than the first will not be designated, as in the other case, simply by ec. To get the true quantity equivalent to e for this second pair of layers we must notice that the e, or the rate at which ξ changes with x, is not the same for different values of x. If we use h to represent the rate at which e changes with x, we may then write hb for the difference between the e for the first pair of layers and the e for the second pair. The amount by which the second layer in this second pair moves farther than the first will be given by $(e + hb)\,c$.

Now the portion of air between either pair of layers is exerting a pressure different from the original general pressure p_0, because the volume of that portion is altered. If we consider the air around the axis of x to be in the shape of a cylinder of unit cross section, the original volume of the portions between the

two layers bounding each portion will be represented by c, and the changes of volume by ec, and $(e + hb)\,c$ respectively. The elasticity E of the air is the ratio of the increase in pressure to the change of volume per unit volume (§ 152). In the case of the first pair of layers, if p_1 represents the pressure between them, and p_0 the original common pressure, we have $E = \dfrac{p_1 - p_0}{\dfrac{ec}{c}}$ or $p_1 - p_0 = Ee$.

Similarly if p_2 represents the pressure between the second pair of layers, we have $p_2 - p_0 = E\,(e + hb)$. The difference of the two pressures p_2 and p_1 is the unbalanced pressure, — in this case, with unit cross section, the resultant force, — on the short column of air of length b between the two pairs of layers. The force acting on this column is therefore $p_2 - p_1 = Ehb$.

If we use ρ to represent the density of the air, the mass m of the column of length b is ρb, and hence the acceleration f of this column, or the ratio of the force to the mass, is

$$f = \frac{p_2 - p_1}{m} = \frac{E}{\rho} h. \tag{83}$$

This is the so-called differential equation of wave transmission, in which the acceleration of each portion of the air is connected with the way in which the displacements of successive portions differ at any instant along the line of transmission.

207. Transmission of Waves in Air. — It is not possible for us to deduce, by elementary methods, an expression for the motion of the layers of air which will be compatible with Equation 83; but by following a procedure essentially similar to Newton's, we may show that if a series of displacements of the sort described by Equation 81a is once set up in the air, the changes of pressure along the column of air are such as to maintain a simple harmonic motion of each layer of air and to transmit simple harmonic waves along the axis. Furthermore, it will result, as a consequence of this investigation, that the velocity of the waves will be equal to the square root of the ratio of the elasticity of the air to its density.

A comparison of the way in which the quantity h has been defined with the way in which acceleration is defined as the rate at which the velocity v changes with the time, while v is defined as the rate at which the displacement ξ changes with the time, will show that h is derived from the changes of ξ with respect to x in the same manner as f is derived from the changes of ξ with respect to t.

We now suppose that at a particular instant the ξ at any point in the air is given by

$$\xi = a \cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

From our study of simple harmonic motion (§ 63) we know that, if x is constant, the acceleration f is given by

$$f = -\frac{4\pi^2}{T^2}\xi;$$

so that by the analogy between the derivation of f and the derivation of h, we may write, if t is constant,

$$h = -\frac{4\pi^2}{\lambda^2}\xi.$$

Using this value for h in Equation 83 we have

$$f = -\frac{4\pi^2 E}{\lambda^2 \rho}\xi. \tag{84}$$

This is the equation defining a simple harmonic motion of period $T = 2\pi\sqrt{-\frac{\xi}{f}}$ (§ 66), or $T = \lambda\sqrt{\frac{\rho}{E}}$. The assumed original series of displacements along the axis will therefore initiate and maintain simple harmonic motions of this period in each layer of air.

Our study of Equation 81a for a simple harmonic wave has shown us (§ 205) that the existence of such motions in the successive layers of air, with phases determined for any particular value of x by the successive values of t, is the distinguishing feature of a train of simple harmonic waves transmitted along the axis with a velocity given by λ/T. The waves resulting from the assumed initial disturbance will therefore travel along the axis with the velocity

$$V = \frac{\lambda}{T} = \sqrt{\frac{E}{\rho}}. \tag{85}$$

208. Fourier's Theorem: Superposition of Small Vibrations. — Though our discussion, of necessity, has been restricted to a special form of wave, yet the result we have reached is of great generality, and applies to sound waves of any form. That this is true depends first upon a theorem in analysis suggested by Daniel Bernoulli, and proved many years afterwards by Fourier. *Fourier's theorem* enables us to describe any initial disturbance of a medium, provided it either is periodic or can be thought of as such, as the sum of a series of simple harmonic disturbances. According to this theorem, any single valued periodic function can be analyzed into a series of simple harmonic functions, whose periods are equal to that of the given function divided by the successive cardinal numbers.

This theorem, therefore, enables us to treat any initial state of the air which is produced by a sounding body as an assemblage of simple harmonic states similar to the one which was assumed in our study of the velocity.

To transfer this analytical resolution of the disturbance to the dynamical problem, we use a principle enunciated by Daniel Bernoulli, called the *principle of the superposition of small vibrations.* This principle may be described as follows:

If a part of an elastic body is set in motion by two or more disturbances reaching it at the same instant, its displacement will

be the resultant of the displacements which those disturbances would impart to it if they were to affect it separately; and the disturbances which it will cause in the parts of the body which lie near it are also the resultants of those which it would have caused if it had been affected by each of these disturbances separately.

For example, if a sound wave is produced on one side of a room, it will pass across the room and affect a portion of the air, in a certain place and at a given instant, in a certain way. The same portion of air will be affected also and differently by another sound wave starting from the other side of the room. If both waves are set up so as to reach the chosen portion of air at the same instant, it will be disturbed by both of them, so that its resultant motion is a combination of the two motions which they will separately produce. The two components of this motion affect the air around this portion independently, and each of the two waves travels on from the point at which they cross each other without being changed in any way. As another example, let us consider a point at the middle of a vibrating string, at which two disturbances from the opposite ends of the string arrive at the same instant. These two disturbances may combine so as to make the resultant disturbance of the point double that which it would acquire from either one of them, or they may combine, if they are of the same magnitude and in opposite directions, so that the resultant disturbance at the point is zero, or the resultant may have any intermediate value. In any case, after the disturbances have passed the point, they will proceed unchanged by any action of the one upon the other.

To apply these principles to the question of the transmission of a general disturbance of any form through an elastic medium, we analyze it by Fourier's theorem into its constituent simple harmonic waves. By the principle of superposition each of these will proceed through the medium as if it alone were present, and the disturbance at any later time will be obtained by superposing them in their new positions. Since the velocity of any simple wave is independent of its peculiar period and wave length, and of its amplitude also, if it is small, and depends only upon the elasticity and density of the medium, all such waves will proceed in the same medium with the same velocity. The constituent waves of which the initial disturbance is composed will therefore all proceed with the same velocity, so that they will always remain in the same relations to one another as at the outset, and the initial disturbance will proceed with the same velocity.

209. The Velocity of Sound. — The conclusion that the velocity of sound is equal to the square root of the ratio of the elasticity of air to its density was the principal result of Newton's study of wave motion. The same general formula for the velocity will

hold for any other medium in which vibrations of a similar sort are set up.

In Newton's time the elasticity of a medium which, like air, obeys Boyle's law, was thought to be equal to its pressure. When Newton used this value of the elasticity to calculate the velocity of sound, he obtained a value for it which was less than that which had been obtained by direct experiment. This discrepancy between theory and experiment was removed many years later by Laplace, who called attention to the fact that the elasticity of air is equal to its pressure only when the temperature of the air is kept constant. Now it is known that a sudden condensation of air will raise its temperature and a sudden rarefaction will lower its temperature. The condensations and rarefactions which constitute a sound wave pass through the air so rapidly that no time is given for the equalization of the differences of temperature which they produce, so that the elasticity which is to be used in the calculation of the velocity of sound is one determined on the condition that heat is neither received nor emitted by the air. The value of this elasticity for air has been measured and found to be equal to its pressure, multiplied by the numerical factor 1.405. With this value of the elasticity, the calculated velocity of sound agreed with the experimental value.

To calculate the velocity of sound, we are given the numerical factor 1.405, the pressure of one atmosphere, 1.0134×10^6 dynes per square centimeter and the density of air under the pressure of one atmosphere and at the temperature of melting ice, 0.001293 grams per cubic centimeter. From these data we calculate that the velocity of sound equals 331.8 meters per second.

If we write the formula for the velocity of sound in the form $V = \sqrt{\frac{\kappa p}{\rho}}$, in which κ represents the numerical factor already referred to, and then substitute for ρ its value m/v obtained from the definition of density, using m to designate a fixed mass of air contained in the variable volume v, we obtain $V = \sqrt{\frac{\kappa}{m} pv}$. Now from Boyle's law we know that the product pv is a constant at a fixed temperature, and therefore the velocity of sound in air will be the same, at the same temperature, whatever be its density. Thus it will be the same at the highest altitudes as it is at the sea level. From the original formula we learn that the velocity of sound in different gases is inversely proportional to the square roots of their densities. Thus in a light gas, like hydrogen or coal gas, the velocity is greater than it is in air.

The velocity of sound was determined experimentally by Mersenne, by noting the time which elapsed between the instant at which the flash of a pistol was seen and the instant at which the

report was heard. By using sounds of different intensities, produced by a cannon and a pistol, Gassendi proved that the velocity of sound is independent of its loudness. That it is also independent of its pitch is shown by the fact that the different tones produced by the instruments of a band or orchestra do not lose their harmonious relations when heard by a distant observer. The statement that the velocity of sound is independent of its loudness is not absolutely correct. Captain Parry relates that when his men were engaged in gun practice in the still air of the Arctic regions, the report of a gun was heard by a very distant observer before he heard the word of command to fire. A study by Earnshaw of the propagation of waves through air has shown that very loud sounds should be propagated with greater velocities.

The mean of three of the best experimental determinations of the velocity of sound at 0° C. gives the value 331.5 meters per second.

210. Transmission of Waves on Strings. — If a string or cord, stretched by a weight or otherwise, is given a smart blow, a portion of it will be distorted, and the distortion will travel along the string with a constant velocity. The displacements of the disturbed parts of the string are transverse to the direction in which the wave travels, but otherwise the transmission of the disturbance seems to be in general similar to the transmission of a short series of waves or a pulse in air.

By following the same general method of argument as that used in the study of the transmission of waves in air, we can show that if the string is thrown into a sinusoidal shape by a train of waves, the tension in it will give rise to forces which act on each short portion or element of the string in such a way as to give it an acceleration proportional to its displacement. Its motion will then be simple harmonic, and will be a consequence of the conditions in the string, so that the assumed mode of motion is dynamically possible. The velocity with which the train of waves will travel along the string will be equal to the square root of the tension in the string divided by the mass of unit length of the string.

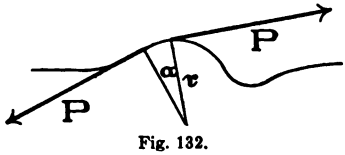


Fig. 132.

Consider any short portion or *element* of the curved part of the string (Fig. 132) of length s. If ρ represents the mass of unit length, the mass of the element is equal to ρs. The element is drawn inward toward the center of curvature by a force which is the resultant of the tensions P which act tangentially upon it at its ends. If the element curves so that the angle between these forces is α,

the resultant (Fig 133) is given by $2P \sin\frac{\alpha}{2} = P\alpha$, if α is very small, so that the sine may be set equal to the arc. The normals to the string at the ends of the element make with each other the angle α, and, if the element is small, they will meet at a point which is the center of curvature of the element. The length r of either of these normals is the radius of curvature of the element, and we have $s = r\alpha$. Dividing the resultant force by the mass of the element to get its acceleration f, and using this equation to eliminate α, we have

Fig. 133.

$$f = \frac{P}{\rho r}. \tag{86}$$

We cannot handle this equation generally by elementary methods, but if we assume that at any instant t the displacement is given by the equation

$$y = a \cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right),$$

we may show, as in the similar case of the train of waves in air, that the acceleration thus represented will be proportional to the displacement, and will thus give rise to simple harmonic motions of the elements of the string, and will transmit waves along the string. It can be proved, in fact, by purely mathematical calculations, that if the elements of the sinusoid represented by the assumed equation for a particular value of t never depart far from a straight line, the displacement and the radius of curvature at any point are connected by the relation $\frac{1}{r} = -\frac{4\pi^2}{\lambda^2} y$.

Setting this value of $1/r$ in the equation for the acceleration, we have

$$f = -\frac{4\pi^2}{\lambda^2}\frac{P}{\rho} y. \tag{87}$$

Strictly speaking, this acceleration is along the normal to the string, but in the assumed conditions it will always be approximately perpendicular to the undisturbed string.

The acceleration of any element of the string is therefore proportional to the displacement of the element, so that the motion of each element is simple harmonic, and the motion, once set up, will be maintained and transmitted in a train of waves. The period of the simple harmonic motion is $T = 2\pi\sqrt{-\frac{y}{f}}$ $= \lambda\sqrt{\frac{\rho}{P}}$; and the velocity with which the wave is transmitted along the string is

$$\frac{\lambda}{T} = \sqrt{\frac{P}{\rho}}. \tag{88}$$

Since this velocity is independent of the peculiar characteristics of the wave, all simple waves will proceed with the same velocity.

By Fourier's theorem (§ 208) any irregular series of displacements of any part of the string can be analyzed into a series of

simple waves, and since these all proceed with the same velocity, the original displacement will proceed with the same velocity, and will always remain similar to itself.

211. **Velocity of Waves on Strings.** — The velocity of any disturbance which is transmitted along a string may be determined by an artifice which is of sufficient interest to receive special discussion. If a disturbance of any curved form is traveling along a string with any velocity, and if the string itself is drawn uniformly in the opposite direction with the same velocity, the form of disturbance will remain fixed in space.

Consider an element of the string (Fig. 134) of length s, and with a radius of curvature r. If ρ represents the mass of unit length of the string, then the mass of the element is given by ρs. As in § 210 we may show that the tension P in the string produces a resultant force on the element directed along the radius of curvature given by Ps/r. Now the elements of the string which are moving through the curved parts of the wave form are constrained to take the curved path at each point by this force, and from the theorem of centripetal force (§ 78) we have $\frac{Ps}{r} = \frac{\rho s v^2}{r}$.

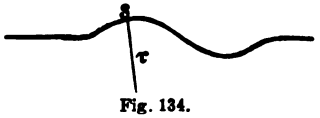


Fig. 134.

From this we obtain immediately $v = \sqrt{\frac{P}{\rho}}$. The interest in this method lies in its not requiring any special hypothesis about the form of the wave or disturbance on the string, except that its curvature is assumed to be everywhere finite.

212. **Transmission of a Disturbance.** — It is a matter of observation that any disturbance set up on a cord or string will travel along it, and that as it passes any point the string at that point comes to rest, and remains at rest thereafter. In fact, any limited wave or train of waves, which passes through or along a body, leaves each point of the body at rest after it has passed. By considering the special case of the string, a general explanation of this important fact may be given, though a full discussion of it is beyond the scope of this book. Consider a string (Fig. 135) of which a portion is distorted from its position of equilibrium by the arrival of a wave form from A. At the front of the wave, at C, the string is initially at rest, and the tension in the string, acting along the radius of curvature, sets the element at C in motion. The motions of the succeeding elements on the wave at any instant are determined by the momenta of the elements and the forces

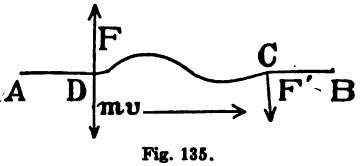


Fig. 135.

due to the tension which act on them. At the rear of the wave, at D, the element is moving in one direction, while the radius of curvature, along which the force acts, is in the opposite direction. The force just suffices to annul the momentum, and to bring the element at D to rest. When the wave has passed the element at D, it does not affect it again.

A generally similar description may be given of the passage of a *pulse* or a terminated disturbance through air. The forces which accelerate the successive layers of air in the pulse arise from pressure differences which vary from layer to layer. At the front of the wave the pressure difference acts upon a layer of air initially at rest and sets it in motion. The motion of the succeeding layers of air, at any instant, is determined by their momenta and the pressure differences which act on them. At the rear of the wave the motion of the layer of air is in one direction, and the pressure difference acts in the opposite direction. The pressure difference just suffices to annul the momentum and to bring the layer to rest.

213. Reflection of Waves. — When a sound wave in air meets a solid obstacle, like a high wall, it is reflected from it; that is, it is turned back in its course and proceeds from then on in a reverse direction. From the analogous reflection which may easily be observed when the waves on water encounter an obstacle, Vitruvius supposed such a reflection to take place in the case of sound, and explained thereby the *echo*. Observation shows that the sound wave does not lose its essential characteristics by reflection.

Reflection of sound occurs whenever the wave meets a surface which separates the medium in which it is traveling from another one. Two cases of reflection must be distinguished, which depend on the relations of the motions of the two media when they are transmitting similar sounds. If the second medium is made up of particles which move over longer distances (Fig. 136, *a*) than those of the first medium do, when they are transmitting similar sounds, and if we consider that portion of the advancing wave in which the particles of the medium are being pushed forward, it is plain that when it reaches the surface of separation the last layer of particles of the first medium will move farther out into the second

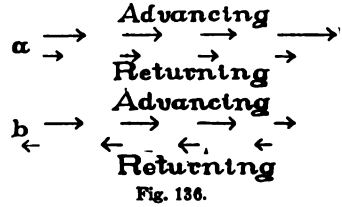


Fig. 136.

than they would have done if the first medium had continued further. With respect to the first medium, this movement of the last layer may be looked on as composed of two movements, one equal to that which the other particles of the first medium have executed, and an additional movement in the same direction. This additional movement is of such a sort as to start a movement in the first medium, proceeding in the opposite direction to that of the original wave. The successive displacements caused by this movement are in the same direction in space, or have the same sign, as the displacements in the original movement by which they were produced. This sort of reflection is therefore called *reflection without change of sign.* Since the sign of the displacement remains the same when the phase remains unchanged, this kind of reflection is also called *reflection without change of phase.*

On the other hand, if the second medium is one whose particles move less freely (Fig. 136, *b*) than those of the first medium, a forward displacement is diminished when it comes to the surface of separation. We may look on the forward displacement of the particles of the first medium at that surface as composed of two displacements, one equal to the original forward displacement, the other less than it in amount and in the opposite direction. This displacement in the opposite direction is of such a sort as to produce a movement in the first medium, proceeding in a direction opposite to that of the original wave. In this case the reflection is said to be *reflection with change of sign.* Since the sign of the displacement changes when the phase is reversed, or increased by the angle π, this kind of reflection is also called *reflection with reversal of phase.*

Similar reflections occur of the waves in strings or other bodies in which such waves may be transmitted. If a heavy string is fastened at one end by a long and light string to some support and is then stretched by a pull by the hand on the other end, a blow on it near the hand will produce a distortion of a part of the string, which will proceed along it as a wave. At the other end the distortion encounters the lighter string, in which freer motion is possible, with the result that the end of the string moves farther out than the intervening parts have done. This motion can be considered as the resultant of two motions, one like that in the rest of the string, the other, an additional motion in the same sense. This additional motion acts as an initial disturbance and is transmitted along the string toward the first end as a wave. The disturbance is reflected without change of sign. If the further end of the string is held fast by a solid clamp, the disturbance, on reaching it, is reduced to nothing. This is equivalent to the motion that would be obtained by the superposition of two motions, one like that on the rest of the string, the other, an equal additional motion

in the opposite sense. This additional motion acts as an initial disturbance, and is transmitted along the string toward the first end as a wave. The disturbance is reflected with change of sign.

EXAMPLES, XIV

1. *To show that the energy of a particle moving with a simple harmonic motion of given period is proportional to the square of the amplitude.*

The velocity of the particle at the middle of its path (§ 64) is $v = \frac{2\pi a}{T}$, and if its mass is m, its kinetic energy at that point is $\frac{1}{2}mv^2 = \frac{2\pi^2 m}{T^2}a^2$. Now this kinetic energy is at that point the total energy of the particle, so that with a given value of T, the energy of the particle is proportional to the square of the amplitude.

If the mass considered is the mass of a continuous medium of density ρ contained in an element of volume ω, its energy is equal to $\frac{2\pi^2}{T^2}\rho\omega a^2$.

2. *To investigate the conditions of reflection.*

Consider a cylinder of length v equal to the distance traversed by the wave in unit time or to the velocity of the sound in the first medium, with its base on the surface separating the two media. When a train of waves is passing through this cylinder the energy of each unit volume is proportional to the density ρ and to the square of the amplitude a_1 of the vibration (Example 1). After reflection this energy is divided between this cylinder, in which is contained the energy of the reflected wave, and a similar cylinder of length v', the velocity of sound in the second medium. Calling the densities of the first and second media ρ and ρ' respectively, and the amplitudes in the reflected and transmitted waves a_2 and a_3 respectively, we may form the equation

$$v\rho a_1^2 = v\rho a_2^2 + v'\rho' a_3^2.$$

The displacements at the surface of separation are mechanically continuous, so that we have also

$$a_1 + a_2 = a_3.$$

From these equations we get

$$a_2 = a_1 \frac{1 - \frac{v'\rho'}{v\rho}}{1 + \frac{v'\rho'}{v\rho}}; \qquad a_3 = \frac{2a_1}{1 + \frac{v'\rho'}{v\rho}}.$$

Now we have $v = \sqrt{\frac{E}{\rho}}$, $v' = \sqrt{\frac{E'}{\rho'}}$, and making the substitutions we have finally

$$a_2 = a_1 \frac{1 - \sqrt{\frac{E'}{E}}\sqrt{\frac{\rho'}{\rho}}}{1 + \sqrt{\frac{E'}{E}}\sqrt{\frac{\rho'}{\rho}}}; \qquad a_3 = \frac{2a_1}{1 + \sqrt{\frac{E'}{E}}\sqrt{\frac{\rho'}{\rho}}}.$$

In the ordinary case of reflection in air by a solid wall, E' and ρ' are both much greater than E and ρ. In that case a_2 is negative and nearly equal to a_1. The reflection takes place with change of sign. The transmitted amplitude a_3 is small.

In the case of reflection in the solid by air, we have a_2 positive and nearly equal to a_1; and a_3 nearly equal to $2a_1$. The reflection takes place without change of sign.

In the case of reflection at a surface separating two gases under the same pressure we have generally $E = E'$, and the amplitudes depend on the densities of the gases. If the second gas is a rare one, like hydrogen or coal gas, $\rho > \rho'$, and a_2 is positive, the reflection is without change of sign. If the second gas is dense, like carbon dioxide, $\rho < \rho'$, and a_2 is negative, the reflection is with change of sign. If $\rho = \rho'$ there is no reflection, and $a_3 = a_1$; the wave goes on without change.

CHAPTER XII

SOUNDING BODIES

214. Resonance. — It was observed by Mersenne that if two strings are tuned to the same pitch, and one of them is sounded, the other will take up a vibration and will also emit a sound. The experiment succeeds especially well if the two strings are mounted on the same sounding board. It was also observed that a string will be set in vibration if another string is vibrating near it which is tuned to the octave or to the twelfth of the tone which it will emit. This general phenomenon of the excitation of vibrations in bodies by the vibrations of other bodies in their presence is called *resonance.*

The phenomenon of resonance is not exhibited by strings only, but by all sorts of sounding bodies. Thus one tuning fork will respond to the vibrations of another which is in unison or in simple harmony with it, and a column of air will respond to the vibrations of a tuning fork or of a string.

To explain resonance we consider the effect of a succession of small impulses applied to a string, which is capable of regular vibration, at intervals which are equal to the natural period of vibration. Such impulses are applied to it by the motions of the air, or of some vibrating body, like the sounding board which supports it, and these are set up by the regular vibrations of the first sounding body. The effect of the first impulse is a slight disturbance of the string. If that impulse were the only one applied to it, it would execute a number of minute vibrations in its own natural period. A second impulse, however, is applied at the end of that period, and therefore just at the time when the string is moving through the position from which it was started by the first impulse. The effect of the second impulse is added to that of the first, and the vibration of the string is increased thereby. As the successive impulses are applied to it, each at the time when its action will increase the vibration, the vibrations become greater and greater until they can often be perceived by the eye and can be heard to give forth a musical tone.

If the natural period of the string is not equal to the interval

between the impulses, the effect of the first impulse will again be to start a small vibration, but each impulse following the first will be applied at a time which departs more and more from that at which the string is in such a position that its vibrations are increased by the impulses. For example, if the string naturally executes 200 complete vibrations in a second, and if the sounding body which sends impulses to it executes 201 complete vibrations in a second, the second impulse is applied one two-hundredth of a second before the first natural vibration of the string would bring it to its most favorable position, the third impulse is applied two two-hundredths of a second too soon, and so on, the successive impulses being applied at times in which the position of the string is less and less favorable. After 100 vibrations have been executed, the impulse which tends to move the string in one sense is applied to it just when it is moving in an opposite sense, and so its effect is to destroy the original vibration. Thus the string will not respond, and resonance will not occur, unless the string has the same period of vibration as the sounding body.

When a string is made to respond to the vibrations of a body which is emitting the octave of the fundamental tone which the string emits, the tone emitted by the string is also the octave. In this case, as was shown by Noble and Pigot, and afterwards by Sauveur, the string is vibrating in two portions. Its middle point is at rest, and each half of it vibrates as if it were an independent string, with its natural period of vibration. When the resonance of a string is produced by a body sounding the twelfth of the fundamental tone of the string, the responding string vibrates in three equal parts. At two points, which divide it into these three equal parts, the string is at rest.

These points at which the string is at rest, when it is emitting a tone which is higher than its fundamental, are called *nodes*. The vibrating portions between the nodes are called *ventral segments* or *loops*. The tones which a string will emit, when it vibrates in parts which are fractional parts of its whole length, are called *harmonics*, *partial tones*, or *overtones*.

An acute ear may detect that when a string is emitting its fundamental tone, it is also emitting some of its harmonics. This observation was made by the investigators whose names have already been mentioned. It shows that the vibration of a string may be much more complicated than the simple vibration which we have assumed so far to be the one executed by it.

In general, what has been said of a string with respect to resonance and with respect to the production of harmonics, not only singly but also in combination with fundamental tones, may be said of all sorts of sounding bodies.

In very many cases a body which is connected with another that is vibrating will be forced into vibrations. It has been shown in Examples, V, 4 that the period of such forced vibrations is that of the applied force, so that the sound produced from these forced vibrations will be of the same pitch as that emitted from the original vibrations. When a natural period of vibration of the body is the same or nearly the same as that of the applied force, the amplitude may become considerable, and the sound emitted very much increased in intensity. Stretched membranes or plates have a large number of natural periods of vibration, and they are accordingly used to increase the sound emitted by many musical instruments.

215. Sounding Bodies. — The phenomena which have now been described directed attention to the mode of vibration of sounding bodies. It was generally assumed, in the earlier study of this subject, that the motion of each part of the body is a simple harmonic motion. Brook Taylor showed that such a motion would be maintained by a string stretched between two fixed points, if its original displacement from its position of equilibrium were such as to make it coincide with one of the arcs of a sinusoid. A string vibrating in this manner will give its fundamental tone.

The simultaneous production of the fundamental tone and of one or more overtones by the same vibrating string, or in general by a sounding body, forces us to admit that the vibrations of the body are not simple harmonic vibrations.

We may, however, by Fourier's theorem (§ 208) analyze the original displacement of the string into a series of simple harmonic displacements, and on the principle of superposition (§ 208) we may assert that whatever is true for one such simple displacement will be true for all of them; so that if a simple harmonic displacement will give rise to a regular vibration of a definite period, each of the displacements into which the given displacement is analyzed will give rise to a vibration of a definite period. These vibrations will combine to produce a general vibration which, while not simple harmonic, is yet periodic, and will be repeated over and over again with a period which is that of the fundamental mode of vibration.

The most direct method of handling the problems presented by sounding bodies would be for us to follow the example set by Taylor, and find an initial set of displacements in the body, consistent with the terminal conditions, which would be such as to initiate and maintain simple harmonic motions in the elements of the body. We may avoid doing this, which necessitates the independent study of each case presented by a sounding body, by the use of an alternative method, for which our previous study has already prepared us.

216. Superposition of Oppositely Moving Waves. — We have seen (§ 210) that the resultant forces acting on each element of a string which is initially shaped in the form of a sinusoid will be such as to set up and maintain simple harmonic motions in each element, and to transmit along the string in the positive direction a train of waves of which the equation is

$$y_1 = a\cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

It is evident on inspection that a similar initial displacement will also set up and transmit along the string in the negative direction a train of waves of which the equation is (§ 205)

$$y_2 = a\cos 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right).$$

Now by the principle of superposition these two trains of waves can exist in the string at once, and we can find the resultant displacement of each element, or the mode of vibration of the string, by adding these displacements.

We have $y_1 = a\cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)$

$$= a\cos 2\pi\frac{x}{\lambda}\cos 2\pi\frac{t}{T} + a\sin 2\pi\frac{x}{\lambda}\sin 2\pi\frac{t}{T},$$

$$y_2 = a\cos 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right)$$

$$= \pm a\cos 2\pi\frac{x}{\lambda}\cos 2\pi\frac{t}{T} \mp a\sin 2\pi\frac{x}{\lambda}\sin 2\pi\frac{t}{T}.$$

Using either the upper or lower signs in the value of y_2 we get

$$y = y_1 + y_2 = 2a\cos 2\pi\frac{x}{\lambda}\cos 2\pi\frac{t}{T}, \tag{89a}$$

or

$$y = y_1 + y_2 = 2a\sin 2\pi\frac{x}{\lambda}\sin 2\pi\frac{t}{T}. \tag{89b}$$

The displacements described by these equations are those of points having simple harmonic motions of the same period, and in the same phase, but differing in amplitude in a way that depends upon the distance of the point from the origin. The general expression for the amplitude, if we take the first value of y, is $2a \cos 2\pi x/\lambda$. The distance x may have such a value, say $x = \lambda/4$, that $\cos 2\pi x/\lambda = \cos \pi/2 = 0$. The point at that distance will thus have zero amplitude, or will always be at rest. If we proceed further along the string by the distance $\lambda/2$, so that $x = 3\lambda/4$, $\cos 2\pi x/\lambda = \cos 3\pi/2 = 0$, we find another point which is always at rest; and similar points at rest will be found at intervals of half a wave length along the whole string. These points are called *nodes*.

At the origin, where $x = 0$, the amplitude of the point is $2a$. At a point on the string at the distance $x = \lambda/2$ from the origin, we have $\cos 2\pi x/\lambda = \cos \pi = -1$, so that the displacement is $-2a$, when that at the origin is $+2a$. At the point distant from this one again by $\lambda/2$, we have $\cos 2\pi x/\lambda = \cos 2\pi = +1$, so that the displacement at that point is $+2a$. The plus and minus values of the displacement $2a$ will alternate along the string at points which are distant from each other by half a wave length. These points are called *antinodes*. The general moving portion of the string between any two nodes is called a *ventral segment*. The disturbances of the string described by our equations are called standing or *stationary waves*.

We may represent the condition of the string when stationary waves exist in it by a diagram (Fig. 137) in which the lines are

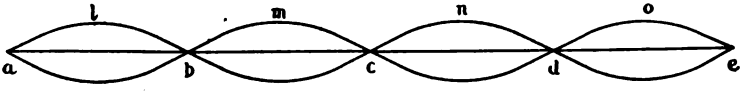


Fig. 137.

drawn to represent the shapes of the string at two instants, those of maximum displacement on either side of the undisturbed string.

The points a, b, c, etc., are the nodes. The points l, m, n, etc., halfway between them, are the antinodes. It will be noticed from the diagram, and may be proved from the way in which the sign of y changes as x passes through a node, that the displacements on the opposite sides of a node have opposite signs. If those on the one side are elevated at a particular instant above the line of the undisturbed string, those on the other side, at the same instant, are correspondingly depressed.

Exactly similar nodes and ventral segments are obtained by the study of the second form of the equation for the displacement. They differ from the first set only in their distances from the origin.

In the same way we may treat the case of two oppositely moving waves in a column of air, and obtain for the resultant displacement in different cases the equations

$$\xi = 2a \cos 2\pi \frac{x}{\lambda} \cos 2\pi \frac{t}{T} \tag{90a}$$

$$\xi = 2a \sin 2\pi \frac{x}{\lambda} \sin 2\pi \frac{t}{T} \cdot \tag{90b}$$

There will be nodes, antinodes, and ventral segments in the column of air, which can be represented by the same diagram as that we have drawn for the waves on a string, if we use the lateral displacement y in the diagram to represent the longitudinal displacement ξ. In this case, since the displacements on opposite sides of a node have opposite signs, the layers of air on the opposite sides of a node are approaching it or moving away from it at the same time and the air around the node undergoes alternate condensations and rarefactions. At the antinodes the layers of air move through the longest paths, but there is no change of density there.

217. Stationary Waves on Strings. — To apply the results now obtained to the case of the vibrations of a string we must first explain how the two oppositely moving trains of waves assumed in the discussion can be set up on the string. To do this we use the conception that each advancing wave, in fact, each displacement of each advancing wave, is reflected at the ends of the string, and returns as a reflected wave, or as a reflected displacement. In the case of the string the reflection is with change of sign at both ends.

A general description may be of service in making it clear that such reflections will account for the regular vibrations of a string

Suppose that a stretched string is plucked near one of its ends. The disturbance produced, which we may call an *elevation*, runs along the string to the other end as a simple wave, where it meets the rigid support in which that end of the string is held. It is therefore reflected with change of sign and comes back as a *depression*. It is again reflected at the end from which it started, and becomes there an elevation, which proceeds along the string again

as has already been described. The string is thus set in vibration. If the part of the string which was first plucked is plucked again and again, at times coinciding with those of the return of the reflected wave to the place of origin, the successive disturbances will be superposed and the vibration will be increased. If the impulses are applied to the string twice as often, so that the advancing wave reaches the other end of the string just as the second impulse is applied, the elevation produced by the second impulse, and the depression due to the first impulse after reflection, will meet in the middle of the string, and the point at which they meet will not be disturbed. It will remain undisturbed as the successive elevations and depressions pass through it, while the parts of the string on either side of it will vibrate. If the impulses are applied three times as often as in the first case, there will be two points, dividing the string into three equal parts, at which elevations and depressions will meet, and which will consequently remain at rest. Similar statements may evidently be made for disturbances applied more frequently still, if only the time between them is some fractional part of the time required by the original disturbance to traverse the length of the string. The points at which the string is at rest are the *nodes*, the parts between them, the *ventral segments*. The disturbance on the string by which its parts are kept in uniform vibratory motion is the standing or *stationary wave*.

The disturbance usually impressed upon a string is not an impulse at one point, but a general displacement of a considerable portion of the string, or even of the whole string at once. By the use of the principle of superposition, we see that the disturbance of each part of the string will be transmitted independently in the way already described for a simple disturbance, and nodes will be produced when the period of the original disturbance has any one of the values already indicated in the preceding discussion.

In the vibrating string, which is clamped between two fixed supports, there is a node at each end. From Fig. 137 we see that these conditions may be satisfied

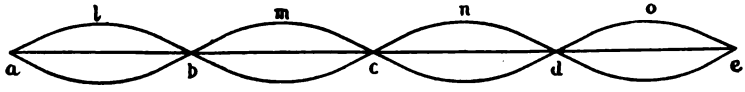


Fig. 137.

when the length of the string is equal to the distance between two nodes, such as *ab*, or to any multiple of that distance. The distance between two nodes is half the wave length of the wave on the string. A string of fixed length *l* may there-

fore be made to vibrate by waves whose lengths are connected with the length of the string by the equations $2l = \lambda_1 = 2\lambda_2 = 3\lambda_3 = \ldots n\lambda_n$. Further, since the velocity of a wave on the string is given (§ 211) by $\sqrt{\frac{P}{\rho}} = \frac{\lambda_n}{T_n} = \frac{2l}{nT_n}$, the frequency, or the reciprocal of the period, is given in general by $\frac{1}{T_n} = \frac{n}{2l}\sqrt{\frac{P}{\rho}}$.

The fundamental tone of the string is given by the vibration corresponding to the longest wave that can subsist in it. The wave length of this fundamental tone is $\lambda_1 = 2l$, or is twice the length of the string, and its frequency is $\frac{1}{T_1} = \frac{1}{2l}\sqrt{\frac{P}{\rho}}$. This formula for the frequency contains the relations between the frequency and the tension, the length of the string and its mass per unit length which were discovered by Mersenne and others (§ 203).

The frequencies given by other values of n are the frequencies of the harmonics which can subsist on the string. Since n may be any number, any one of the harmonics may be set up on the string. Furthermore, by the principle of superposition the various possible harmonics may subsist on the string together, and the string may give out its fundamental and all the harmonics. When the string is bowed or plucked the vibration is generally at first very complex, but the internal friction and the loss of energy to the air soon cut down the vibrations which give the higher harmonics and the sound becomes more and more simple. From the diagram (Fig. 137) it can be seen at once that, when $l = \frac{n\lambda_n}{2}$, there are in all $n + 1$ nodes on the string, and n ventral segments.

The formula for the displacement which embodies the conditions of vibration which we have now deduced is in the general case, when $l = \frac{n\lambda_n}{2}$ and $\frac{1}{T_n} = \frac{n}{2l}\sqrt{\frac{P}{\rho}}$,

$$y = 2a \sin 2\pi \frac{x}{\lambda_n} \sin 2\pi \frac{t}{T_n},$$

in which the values of λ_n and T_n may be substituted to show how the displacement depends on the length of the string and the number of ventral segments into which it is divided.

218. Stationary Waves in Pipes. — The waves in a pipe are due to the vibrations of the column of air in the pipe. In the case of the *open pipe*, in which the column of air is in free contact with the external air at both its ends, the friction of the particles in the pipe against the walls puts the air in the column under so much restraint that the reflection at the ends is at the surface of a medium which is more free, and is therefore reflection without change of sign. The layers of air at the ends of the column have as large an amplitude of vibration as any part of the column.

From Fig. 137 we see that these conditions may be satisfied when the length of the pipe is equal to the distance between two antinodes, such as lm, or to any multiple of that distance. The distance between two antinodes is half the wave length of the wave in air. A column of air in a pipe of fixed length l may there-

fore be made to vibrate by waves whose lengths are connected with the length of the pipe by the equations

$$2l = \lambda_1 = 2\lambda_2 = 3\lambda_3 = \cdots n\lambda_n.$$

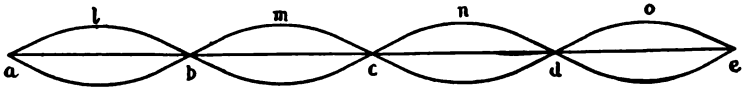


Fig. 137

Further, since the velocity of a wave in air is given (§ 207) by $\sqrt{\frac{E}{\rho}} = \frac{\lambda_n}{T_n} = \frac{2l}{nT_n}$, the frequency is given in general by $\frac{1}{T_n} = \frac{n}{2l}\sqrt{\frac{E}{\rho}}$.

The fundamental tone of the pipe is given by the vibration whose wave length is $\lambda_1 = 2l$, and whose frequency is $\frac{1}{T_1} = \frac{1}{2l}\sqrt{\frac{E}{\rho}}$. The shorter waves, corresponding to the other values of n, give the harmonics. Since n may have any value, any one of the harmonics can be emitted by an open pipe; and more generally, on the principle of superposition, the open pipe can emit a fundamental tone and all its harmonics.

Inspection of the diagram will show that, when $l = \frac{n\lambda_n}{2}$, there are $n + 1$ antinodes, and n nodes in an open pipe.

The formula for the displacement, which embodies the conditions of vibration which we have now deduced, is in the general case, when $l = \frac{n\lambda_n}{2}$ and $\frac{1}{T_n} = \frac{n}{2l}\sqrt{\frac{E}{\rho}}$,

$$y = 2a\cos 2\pi\frac{x}{\lambda_n}\cos 2\pi\frac{t}{T_n},$$

in which the values of λ_n and T_n may be substituted, to show how the displacement depends on the length of the pipe and on the number of half wave lengths contained in it.

In the case of the closed or *stopped pipe*, reflection still occurs at the open end without change of sign, but at the stopped end the layer of air next the solid boundary is necessarily at rest, and reflection occurs there with change of sign. There must therefore be an antinode at one end and a node at the other.

From Fig. 137 we see that these conditions may be satisfied when the length of the pipe is equal to the distance between an antinode and a node, such as *lb*, or to any odd multiple of that distance. The distance from an antinode to a node is a quarter of the wave length of the wave in air. A column of air in a pipe of fixed length l may therefore be made to vibrate by waves whose lengths are connected with the length of the pipe by the equations $4l = \lambda_1 = 3\lambda_3 = 5\lambda_5 = \ldots (2n-1)\lambda_{2n-1}$. As in the case of the open pipe, the frequency is given in the general case by $\frac{1}{T_{2n-1}} = \frac{2n-1}{4l}\sqrt{\frac{E}{\rho}}$.

The fundamental tone is given by the vibration whose wave length is $\lambda_1 = 4l$, and whose frequency is $\frac{1}{T_1} = \frac{1}{4l}\sqrt{\frac{E}{\rho}}$. The frequency of the fundamental from

a stopped pipe is half that of the fundamental from an open pipe of the same length, and the tone is consequently an octave lower.

The shorter waves, corresponding to the other values of n, give the harmonics. Since their frequencies are the odd multiples of the frequency of the fundamental, the stopped pipe emits only the odd harmonics.

From the diagram (Fig. 137) it can be seen at once that, when $l = \frac{(2n-1)}{4}\lambda_{2n-1}$, there are in all n antinodes, and n nodes in a stopped pipe.

The formula for the displacement, which embodies the conditions of vibration which we have now deduced, is in the general case, when $l = \frac{2n-1}{4}\lambda_{2n-1}$, and $\frac{1}{T_{2n-1}} = \frac{2n-1}{4l}\sqrt{\frac{E}{\rho}}$,

$$y = 2a\cos 2\pi\frac{x}{\lambda_{2n-1}}\cos 2\pi\frac{t}{T_{2n-1}},$$

in which the values of λ_{2n-1} and T_{2n-1} may be substituted.

219. The Production of Vibrations. — In order to develop stationary waves, it is not necessary to apply to the body impulses which are properly timed. It is sufficient to apply a large number of impulses, among which there will be many which occur at the proper times. As has already been explained in § 214, the body will respond to those impulses, and will not be affected by the others. For example, when a bow is drawn across a violin string, the impulses which it applies to the string have no regularity, yet the string responds to those which occur at intervals corresponding to its fundamental period, and also to those occurring at intervals corresponding to the periods of some of its harmonics. Those harmonics cannot occur which would establish nodes or fixed points in that part of the string which is kept moving by the bow.

Similarly the confused noise made by blowing upon the lip of an organ pipe does not itself produce the sensation of a clear tone, but among the many vibrations contained in it, those corresponding to the fundamental of the pipe and its harmonics are selected and enhanced by the resonance of the column of air in the pipe.

220. Longitudinal Vibrations of Rods. — A long rod of wood, glass, or metal, if clamped in the middle and set in vibration by stroking it along its length, will execute vibrations and will maintain stationary waves similar to those of the open pipe. As in the case of the air in the pipe, the length of the wave in the material of which the rod is composed, whose period is that of the emitted tone, is twice the length of the rod.

An arrangement by which the velocity of sound in a solid may be determined, depending upon this fact, was invented by Kundt. It consists (Fig. 138) of

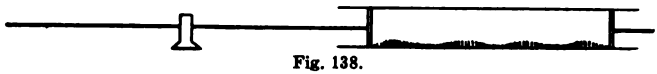


Fig. 138.

a rod of the material under investigation, clamped in the middle, and with one end inserted lightly through a cork, which closes one end of a glass tube. Fine light powder is scattered within the tube, and its other end is closed with a loosely fitting cork, that can be moved up and down in it. When the rod is made to vibrate longitudinally, vibrations of the same period are impressed upon the column of air in the tube, and by setting the movable cork at the right place, they may be made to set up standing waves in the tube. These waves stir up the powder which has been scattered in the tube, and arrange it in a regular pattern, from which the positions of the nodes may be determined, and so the length of the wave in air, whose period is that of the tone emitted by the vibrating rod. Since the length of the corresponding wave in the rod is twice the length of the rod, the ratio between twice the length of the rod and the length of the wave in air is the ratio of the velocity of sound in the rod to the velocity of sound in air. If some gas other than air is used in the tube, the wave length determined in it, compared with the wave length produced in air by similar vibrations of the rod, affords a means of determining the velocity of sound in the gas and so of testing Newton's formula.

221. Transverse Vibrations of Rods. — If a rod is clamped at one end, it may be made to execute transverse vibrations, which will, if sufficiently rapid, give rise to a musical tone. When the rod is emitting its fundamental tone, there is a node at the end at which it is fixed. The first harmonic which is developed in such a rod is about three octaves higher than the fundamental.

When a rod which is free at both ends vibrates transversely, it develops two nodes, which are distant from the ends about two-ninths the length of the rod. The rod may be supported at the points where these nodes occur without interfering with its vibrations. If the rod is bent in the middle, the two nodes approach each other, until, when the two halves are parallel, the nodes are very near together. The *tuning fork* is an example of a rod bent in this manner. As the ends of the fork swing toward each other, the middle of the fork, to which is attached the stem or handle, moves downward. As the ends move apart, the middle of the fork moves upward. Thus, if the fork is sounding, and the handle is brought down on a table, it will tap the table top at regular intervals. The vibrations thus imparted to the table are communicated from it to the air, and thus the volume of sound emitted by the fork is considerably increased.

222. Vibrations of Plates. — Stationary waves may also be developed in plates of metal or glass, which are firmly clamped at some one point and are bowed or stroked at some point on the edge. The nodes of the waves thus produced are detected by sprinkling a little sand over the plate, which is thrown away from the vibrating parts of the plate and collects at the nodes. This method of observation was introduced by Chladni, and the figures obtained are known as Chladni's figures.

223. Quality of Sounds. — When two musical instruments are sounding the same tone, the sounds which they emit, however exactly they may agree in pitch, differ entirely in *quality*. We may trace this difference in quality to the other tones beside the fundamental which the bodies are producing. To take the simplest case, tones of the same pitch coming from an open and a closed organ pipe are perceptibly different in quality, the tone from the open pipe being fuller and richer than the other. This difference is due to the circumstance already discussed (§ 218), that the column of air in the open pipe maintains stationary waves corresponding to all the harmonics of the fundamental, while that in the closed pipe maintains only those corresponding to the uneven harmonics. When a pipe is sounded these harmonics are emitted as well as the fundamental, and the quality of the tone is determined by their presence or absence in the sound and by their relative intensities. In the case of strings all harmonics may be present, and further additional tones occur, which are produced by very rapid vibrations which differ considerably with the material of which the string is made. It is in general true of all wind instruments also that the tones emitted by them are distinguished as respects their quality not only by the relative intensity of their various harmonics, but also by characteristic tones, or even by characteristic noises, which depend upon the material and the construction of the tube in which the air column is contained, and also upon the way in which the original vibrations are produced.

224. Observation of Harmonics. — As a means of observing the presence of a harmonic we use an instrument designed by Koenig, called the *manometric capsule*. This consists of a small box divided into two chambers by a thin flexible membrane. One of these chambers is kept filled with illuminating gas, which is burned in a small jet at the end of a tube projecting from the chamber. The other chamber is connected by a tube to the source of the sound. When the sound is given forth, the gas jet is viewed in a revolving

mirror. The vibrations of the membrane due to the sound produce changes in the height of the flame, and the band of light which is seen in the mirror appears serrated at the top. If the tube of the manometric capsule is introduced into the side of an organ pipe, at a place where a node exists, the alternations of rarefaction and condensation which occur at a node will be demonstrated by a marked serration of the band of light. If, on the other hand, the tube is introduced at the middle of a ventral segment, where there are no changes of density, the serrations of the band will not appear.

The harmonics of pipes may be produced with but little admixture of the lower tones by changing the intensity of the blowing and often by a manipulation of the mouthpiece. The existence of the nodes and ventral segments corresponding to the harmonics may be demonstrated by the manometric capsule.

By touching a string at the middle and bowing it, it can be made to emit the octave of its fundamental. When it is sounding the octave, a light rider of paper may be placed at the middle point and will remain there undisturbed, whereas a rider anywhere else on the string will be thrown off. Higher harmonics may be produced in a similar way by touching the string at other points, and the existence of the nodes and ventral segments corresponding to them may also be demonstrated by the use of riders.

The most complete way by which to study a composite tone is by the use of resonators. The *resonator* is a hollow metallic sphere with a circular opening on one side, and on the other a short projecting tube, to which the ear may be placed, or on which the tube of the manometric capsule may be attached. The air within the sphere, considered as a sounding body, has a fundamental mode of vibration and but very few and feeble harmonic vibrations. If this resonator is in the presence of a body which is emitting the fundamental tone of the resonator, the air in it will be set in vibration, and its vibration will be indicated by the manometric capsule. A series of such resonators may be made, tuned to the different tones of the scale, or to the successive harmonics of a fundamental tone, and can then be used to investigate the different tones emitted at the same time by a sounding body. It was by the use of such an instrument, called an *analyzer*, that Helmholtz demonstrated the relation between the quality of a tone and the harmonics present in it.

225. Beats and Resultant Tones. — When two tuning forks, or two pipes, which have nearly the same pitch, are sounded together,

the sound heard varies in intensity periodically. The complete change of intensity which occurs, from the greatest intensity heard through the least to the greatest again, is called a *beat*. Beats may be explained in the following way: The two sounding bodies are vibrating in periods which are nearly, but not quite, the same. At a certain instant they are moving in such a way as to affect the air around them in a similar manner. The disturbance in the air is then twice as great as that which either of them would produce if it were sounding alone. As they continue to vibrate, their motions become less and less similar, until, after a certain time has elapsed, they are exactly dissimilar and are affecting the air around them oppositely. The sound then has its least intensity. From that time on the sound will gradually increase in intensity, until one of the bodies has gained a whole vibration on the other, when the sound will again have its greatest intensity. Since the time between two instants of greatest intensity is that required for one body to execute one vibration more than the other, the number of beats heard in one second will be equal to the difference between the frequencies of the two bodies.

Koenig has shown that beats may also be heard when two bodies are sounding together whose pitch is very different, provided that the two tones are nearly an octave apart, or are nearly in some other harmonious relation.

When two tones are sounded together whose vibration numbers differ considerably, a third tone is often heard, called the *resultant tone*. The frequency of this tone is equal to the difference between the frequencies of the constituent tones. Koenig's observations make it probable that resultant tones are caused by the regular beats that occur when the two primary tones are sounded. Other resultant tones are heard when two primary tones are sounded strongly, which Helmholtz explained on the hypothesis that the movements of the air set up by the sounding bodies are so great that the simple law of elasticity is no longer applicable.

226. Harmony. — It is interesting to observe that those tones whose combination is most pleasing and gives most fully the sense of harmony are those whose frequencies are in the simplest proportion to each other. Thus, next to the combination of two tones which are in unison, or whose frequencies are the same, the most perfect harmony is obtained when the octave is sounded with the fundamental; that is, by the combination of two tones whose frequencies are as 2 to 1. The next most harmonious combination is

that of the fifth with the fundamental, whose frequencies are as 3 to 2. The combinations of the fourth with the fundamental, and of the third with the fundamental, whose frequencies are as 4 to 3, and as 5 to 4, respectively, are also harmonious. On the other hand, the combination of the seventh with the octave, whose frequencies are as 15 to 16, is not harmonious. In this case beats are heard. It seems probable that discordant combinations or discords occur when beats are produced which lie between certain limits. Tones which produce fewer beats than 10 per second, or more beats than 70 per second, are not discordant, but if the number of beats produced lies between those limits, the tones are discordant.

227. Absolute Frequency. — So far we have considered mainly the relative number of vibrations executed by a sounding body, or producing a sound, in terms of the vibrations of some body taken as a standard. It is, however, a matter of interest to determine the absolute frequency which corresponds to a certain pitch. The determination of this number may be made in several ways.

It has already been explained how Mersenne, by using a string so long that its vibrations could be counted, and comparing its length with that of a shorter string under the same tension and emitting a standard tone, was able to determine the number of vibrations corresponding to that tone.

Sauveur used the beats produced by two organ pipes, whose frequencies were in a known ratio, to determine their absolute frequencies. The pipes which he used gave the fundamental and the seventh of the next lower octave. Their frequencies were therefore in the ratio of 16 to 15. In the sound which was heard when they were sounded together there were six beats per second, so that the difference of their frequencies was 6. From these two relations it follows that the frequencies of the two tones were 90 and 96.

Savart used a toothed wheel, which was rotated by a mechanism, so constructed that its rate of rotation could be controlled and that the number of rotations per second could be counted. When the wheel was in rotation and a flexible card brought up against the teeth, a musical tone was emitted, and by regulation of the rate of rotation this tone could be brought in harmony with that of a standard pipe or fork. The number of vibrations in the standard tone was then determined from the number of rotations made by the wheel in a second, and the number of teeth on the wheel.

Duhamel set a light pointer on the end of a tuning fork, and arranged it so that the fork when sounding was carried along over a plate of glass covered with a thin coating of lampblack. The pointer traced out a sinuous line on the blackened surface, and the number of vibrations executed by the fork was equal to the number of sinuosities in this line. By means of an additional mechanism, a timekeeper was made to record equal intervals of time on the same blackened surface, and by counting the number of sinuosities lying between two marks made by the timekeeper, the frequency of the fork was determined.

The frequency of a tone may also be determined by means of the siren. The *siren* is an instrument which produces a sound by means of the alternate emission and suppression of puffs of air at regular intervals. It consists of a box or air chest with a flat top, in which are pierced a number of holes set uniformly around a circle. A flat disk is mounted on an axle, which turns in a support set at the center of this circle, and the disk is set as near the top of the box as it can be without touching it. A set of holes is made in the disk, which correspond in number and position to those in the lid of the box. Sometimes, by setting the holes obliquely, the air which is forced out of the box may be made to turn the disk, but it is best to turn the disk by some additional mechanism. A counter is provided by means of which the rate of rotation of the disk may be determined. When the disk is turned and air is forced into the box, a puff of air comes out of each hole when the holes in the disk stand directly over those in the box. When the disk turns on so that the holes in the box are covered, the air is shut off. The number of puffs emitted during one rotation of the disk is equal to the number of holes in the disk. To determine the frequency of a given tone, the disk is turned until the tone emitted by the instrument is in unison with the one whose frequency is desired. When this condition is obtained, the rotation of the disk is maintained uniform, and the number of its rotations per second is determined by the counter. From this number and the number of holes in the disk, the frequency of the tone may be obtained.

BOOK IV

HEAT

CHAPTER XIII

TEMPERATURE AND HEAT

228. Sensation of Temperature. — The sensation by which we distinguish between hot and cold bodies is, and must always have been, a perfectly familiar one. The sense by which we perceive it may be called the *temperature sense.* The sensations given by this sense depend upon so many conditions that they are utterly untrustworthy as a measure of the temperature of the body which we examine by it. Thus the same mass of water will appear to the hand either warm or cold, according as the hand has previously been immersed in very cold or in very hot water; and two bodies of different material, like wood and iron, which have been exposed to the same conditions, and which all physical tests show us must be at the same temperature, will not appear equally hot or cold when tested by the hand.

It has always been customary to speak of the cause of the sensation felt, when we touch a hot body, as heat, and to assign the different sensations given us by the body in different circumstances to the presence in it of more or less heat. Similarly the sensation experienced when we touch a cold body has often been assigned to the presence in the body of something called cold. The distinction between the sensation itself and its cause is often not sharply drawn, and a great deal of confusion exists in the early work on the subject between the two ideas of temperature and quantity of heat. Our sensations immediately recognize temperature and changes of temperature, and attention was first turned to the study of those physical relations of bodies which are connected with their temperature.

Our sensations are so uncertain in their estimate of the temperature of bodies that it is of first importance to obtain an instrument which will indicate temperature. Such an instrument is called a *thermometer.*

229. Galileo's Thermometer. — The first thermometer was constructed by Galileo. It consisted of a small glass globe fitted with a long tube (Fig. 139). The open end of the tube was inserted in water, and the globe was heated until some of the air was expelled from it. When it was allowed to cool, the pressure of the external air was greater than that of the air in the globe, and so a column of water was forced up the tube, until equilibrium was established. The air thus shut off in the globe formed what we may call the *thermometric body*, and the top of the water column determined its volume. In using this instrument, as in using any thermometer, it was assumed, on the basis of general experience, that when two bodies whose temperatures are different are brought into each other's presence, their temperatures will finally become equal. This equalization of temperature is brought about by a lowering of the temperature of the hotter body and a raising of the temperature of the colder body. It was also assumed that a change in the temperature of a body is accompanied by a change in its volume, and in particular that the volume of air increases as its temperature rises.

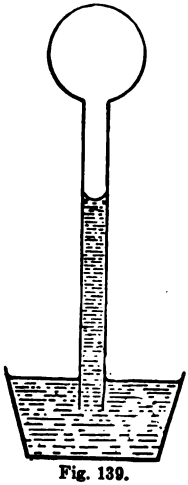


Fig. 139.

With these fundamental principles the mode of operation of Galileo's thermometer is evident. When it is kept in a room at the ordinary temperature, the top of the water column will stand at a certain point in the tube, which may be marked. If it is then transferred to a hotter room, or if a hot body is placed around the globe, the equalization of temperature already described will occur, and the volume of the air in the instrument will increase. The distance through which the top of the water column is forced down the tube is a measure of the change of temperature. For various reasons, the principal one of which is the irregularity introduced by changes in the external atmospheric pressure, Galileo's thermometer will not give consistent indications of temperature. It was very soon superseded by instruments constructed on a different plan.

230. Thermometers with One Standard Temperature. — Stimulated, no doubt, by this invention of Galileo, a body of physicists, resident at Florence and united in a club called the Accademia del Cimento (1657–1667), undertook the construction of thermometers

which should furnish satisfactory measures of temperature. The instrument they made was similar in form to the ordinary thermometers now in use. That is, it consisted of a glass bulb joined to a graduated tube, the bulb and part of the tube being filled with a liquid. The peculiarity of their instrument consisted in this, that they attempted to graduate the tube in such a way that the volume between two marks of graduation was a fixed fractional part, generally one-thousandth, of the volume of the bulb. The instrument thus made was filled with alcohol, so that the top of the column stood opposite a chosen mark on the scale when the instrument was exposed to some standard temperature. The temperature chosen as standard was that of the air during the first light frosts at the beginning of the winter. The first of these instruments were open at the top, but as this interfered with their permanence and with their transportability, they were afterwards closed. This plan of constructing thermometers did not succeed in furnishing instruments which would give similar indications at different temperatures, and notwithstanding the partial success which was obtained much later by Réaumur in the construction of thermometers on the same plan, it has long ago been abandoned.

231. **Thermometers with Two Standard Temperatures.** — The method of graduating thermometers which is now universally used was described by Dalencé in 1688, and an instrument made on that plan was constructed by Newton. In it no attempt is made to establish any fixed relation between the volumes indicated by the graduation and the volume of the bulb. Only so much choice is exercised of the relative volumes of the tube and the bulb as will insure that the thermometer can be used throughout the temperature range for which it is intended. Newton used linseed oil as the thermometric substance. To graduate the instrument, he placed it first in a mixture of ice and water, the temperature of which was known to remain constant. He made a mark on the tube at the point indicated by the top of the liquid column, after the instrument had stood in the ice for some time and showed no signs of any further change. This point indicated one standard temperature. He then placed the bulb of the instrument under his armpit, and after the column again became stationary, he made another mark on the tube opposite the end of the column. The temperature of the human body was known to be very nearly constant, and the temperature thus obtained was therefore taken by Newton as a second standard temperature. The distance be-

tween the two standard marks thus obtained was divided into twelve equal parts, and the graduation thus established was extended above and below the standard marks. Newton did not choose the best thermometric substance that can be used, or the most suitable standard temperatures, but the method which he employed was essentially correct. It is easy to see that if two instruments in which the same thermometric substance is used are graduated in this manner, they will not only indicate the standard temperatures under the same conditions, but will agree in their indications, to whatever temperature they are exposed.

232. Graduation of Thermometers. — The use of mercury as a thermometric substance was introduced by Fahrenheit, to whom we owe the first thermometers which compare with those now made in the accuracy of their indications. Fahrenheit (1714) used the temperature of melting ice as one of the standard temperatures, and probably the temperature of boiling water under a standard pressure as the other. It was subsequently discovered that the temperature of boiling water not only depends upon the atmospheric pressure, but also upon the material of the vessel in which the water is boiling. Cavendish therefore proposed to use as the second standard temperature the temperature of steam over boiling water, which he showed to be independent of the material of the vessel. It depends upon the atmospheric pressure in a way which has been carefully determined by experiment, so that a correction can be made to standard pressure. These two temperatures, the temperature of melting ice and the temperature of steam over boiling water at a standard pressure, are those now universally adopted as the two standard temperatures.

Fahrenheit assigned the number 32 to the first standard mark, and the number 212 to the second standard mark. He therefore divided the distance between the two marks, or, as we may call them, the freezing point and the boiling point, into 180 parts or degrees. When this scale is extended below the freezing point, the zero marks a temperature which is very nearly that of a mixture of pounded ice and salt.

A change in the number of degrees between the freezing point and the boiling point was recommended by Linnæus and carried out by Celsius. In the thermometers of Celsius there were 100 degrees between the two standard temperatures. The temperature of the freezing point was marked 0, so that the temperature of the boiling point was 100. The scale thus constructed is now

generally called the Centigrade scale. It is the one universally used in physical investigations, though the Fahrenheit scale is still used in England and America by meteorologists.

233. Temperature Scales. — It should be observed that the scales of temperature which have been described are purely arbitrary ones. The standard temperatures are arbitrarily chosen, and the change of temperature which is called a degree is determined by an arbitrarily chosen change of volume of an arbitrarily chosen substance. There is nothing which tells us that the change of temperature which causes this standard change of volume in one part of the scale is the same as the change of temperature which will cause the same change of volume in another part of the scale, it being understood that by change of temperature in this statement is meant a change measured by a change in the fundamental physical condition of the body, whatever it may be, which is the true measure of its temperature. These thermometric scales, therefore, do not furnish an absolute measure of temperature, and strictly speaking only indicate differences of temperature.

Many other arrangements are used for the measurement of temperatures and temperature differences. They depend for their operation upon some relation between some physical property of a body, which can be measured, and its temperature. Their indications are usually compared with those of a thermometer, and are thus expressed in degrees.

234. Melting Temperatures. — Even with their imperfect thermometers the early observers made several important discoveries in connection with the subject of temperature. The Accademia del Cimento tried the experiment of immersing a vessel filled with ice in a large mass of hot water, and observing the temperature of the ice with one of their thermometers. They expected to find that the temperature of the ice would fall, but found, in fact, that it remained constant until the ice was melted. By trial with other masses of ice they found that ice always melts at appreciably the same temperature. This temperature is called the melting point of ice. Their observation was confirmed by other observers, and it was shown further that many other substances possess definite melting points, which are characteristic of the substances. By selecting a set of such substances, melting at different temperatures, a series of definite temperatures may be determined, which are independent of the construction of any particular instrument,

and, so far as we know, will be the same everywhere. The first scale of this sort was established by Newton.

235. Boiling Temperatures. — Hooke discovered that when water is boiled its temperature is always approximately constant. This temperature, called the boiling point of water, is not so independent of external conditions as the freezing point is. In particular, it depends upon the pressure in the vessel in which the water is boiled. This had been proved some time before Hooke's discovery by Boyle, who placed a vessel of water which was hot, but not boiling, in the receiver of his air pump, and exhausted the air from around it. When the exhaustion had reached a certain point, the water began to boil, and it could be made to boil again and again by exhausting the receiver still further, although it was continually cooling. On the other hand, if water is enclosed in a tight vessel, like a boiler, under high pressure, it will not boil unless its temperature is raised far above its boiling point in an open vessel.

When water is boiled in an open vessel, it is under atmospheric pressure, and this changes from time to time, so that the boiling point of water, when tested by an accurate thermometer, will not appear to be the same at all times. It will be the same, however, if examined at times when the atmospheric pressure is the same. In order to use the boiling point as a standard temperature, the standard pressure of one atmosphere has been selected as the one at which the boiling point shall be standard.

Many other liquids have been found to have equally definite boiling points, which depend upon the pressure in a way generally similar to that described in the case of water.

236. Freezing Mixtures. — The Accademia del Cimento discovered several pairs of substances which, when mixed with each other, would produce very low temperatures. Such mixtures were called *freezing mixtures*, because the temperatures which they produced were so low that other bodies could be frozen by means of them. A mixture of ice and salt is a familiar example. If a quantity of broken ice, at a temperature below the freezing point, while it is therefore a dry solid, is mixed with salt, the temperature of the mixture falls until it reaches 0° Fahrenheit, or about −18° Centigrade. Of course it cools the bodies around it, and thus may be used to freeze a liquid brought in contact with it.

It was noticed by Boyle, who devoted considerable attention to freezing mixtures, that the solid bodies which were brought to-

gether in the mixture always melted, or that, if they were melting already, they melted faster after being mixed.

237. Freezing Temperatures. — When water is exposed in a vessel, for a sufficient time, to a temperature which is below its melting point, it will gradually freeze. While freezing, its temperature remains constant at the melting point, so that the melting and freezing temperatures are the same. This fact was not recognized by the earliest observers, because of the way in which the temperatures of different parts of a mass of water differ on account of their differences in density, but it was easily observed when small quantities of water were rapidly frozen by means of freezing mixtures. The same general statement holds true for other bodies which have definite melting points, that their melting points and freezing points are the same.

238. Supercooling. — An apparent exception to this rule was discovered by Fahrenheit. He showed that if a small quantity of water was first boiled, so as to expel the air from it, and then allowed to cool slowly in a smooth glass vessel, its temperature might fall several degrees below the freezing point without its freezing. The water in this condition is said to be supercooled. If supercooled water is suddenly agitated, or if a grain of sand, or better still a crystal of ice, is dropped in it, it will at once begin to freeze. Freezing goes on in this case much more rapidly than when it begins at the freezing temperature. At the same time the temperature rises to the normal freezing point. Very many liquids may be supercooled in a similar way, and exhibit similar phenomena when they freeze.

239. Freezing and Boiling Temperatures of Solutions. — The freezing point of a solution is generally lower, and the boiling point higher, than those of the pure solvent. If the substance dissolved, or solute, is one of those for which the osmotic pressure (§ 199) is such as to indicate that its molecules are not dissociated, and if the solution is always very dilute, the amount by which the freezing point is lowered is proportional to the strength of the solution. If different solutes of this sort are dissolved in equal masses of the same solvent in amounts proportional to their molecular weights, the freezing points of the solutions are all lowered by the same amount. For such solutions the amount by which the freezing point is lowered is characteristic of the solvent and is independent of the nature of the solute. For example, if cane sugar, dextrose, mannite, and other similar organic compounds are dissolved in

1000 grams of water, in amounts in grams equal to their respective molecular weights, the common depression of the freezing points of the solutions thus formed is 1.855° C. This is called the *molecular depression* of the freezing point of water. The molecular depression of the freezing point of benzine is 5.1° C., of acetic acid 3.9° C.

Similar laws apply to the elevation of the boiling point. It is, for dilute solutions, proportional to the strength of the solution, and is the same, for the same solvent, when the solutes are dissolved in amounts proportional to their molecular weights. The *molecular elevation* of the boiling point is for water 5.2° C.

Most salts and many other substances occasion larger depressions of the freezing point and elevations of the boiling point than those here described. This fact is explained by supposing that the solute undergoes dissociation (§ 199) when it is dissolved.

240. Change of Volume on Freezing. — When a liquid freezes, there generally occurs a rearrangement of its parts, such that the density of the solid formed is different from that of the liquid. Galileo noticed that this is true in the case of water and ice, and showed, from the fact that ice floats in water, that the density of the ice is less than that of the water. The relative density of ice to water is about as 0.918 to 1. Metals like bismuth or type metal, with which sharp castings can be made, agree with water in having the density of the solid state less than that of the liquid. In most cases the change is in the opposite sense, and the density of the solid is greater than that of the liquid.

Heat

241. The Temperature of Mixtures. — Taylor, and afterwards Richmann (1711–1753), tried the experiment of mixing quantities of water together whose temperatures were different, and observing the resulting temperature of the mixture. They found that, when the two quantities of water were equal, the resulting temperature was the mean of the original temperatures. When the quantities of water were not equal, this was not the case. The resulting temperature was found to be given by the following rule: Multiply the mass of each portion of water by its temperature, add the products, and divide the sum by the sum of the masses; the quotient is the resulting temperature. This rule is known as *Richmann's rule.*

If we designate by m_1 and m_2 the masses of water, by θ_1 and θ_2

their respective initial temperatures, and by θ the final common temperature, Richmann's rule is expressed by the formula

$$\theta = \frac{m_1\theta_1 + m_2\theta_2}{m_1 + m_2}. \tag{91}$$

By experiments with other liquids, Richmann's rule was found to be generally applicable to all cases in which two portions of the same substance are mixed.

242. Heat as a Measurable Quantity. — To explain any ordinary changes of temperature in a body, we assume that they are due to the passage of something, called *heat*, into the body or out of it. When the temperature of the body rises, we assume that it receives heat; when its temperature falls, that it gives out heat. The amount of heat which passes obviously depends on the mass of the body and on its change of temperature. By a study of Richmann's rule, we find a basis for the assumption that it is simply proportional to the mass and to the change of temperature. If we write the formula embodying Richmann's rule in the form

$$m_1(\theta_1 - \theta) = m_2(\theta - \theta_2),$$

we have an equation between two similar quantities, one of which is equal to the mass of the hotter portion multiplied by its fall in temperature, the other, to the mass of the colder portion multiplied by its rise in temperature.

Now we assume that these changes of temperature have taken place because of the passage of a quantity of heat from the hotter to the colder portion, so that, if this quantity of heat, considered with respect to the hotter portion, is proportional to $m_1(\theta_1 - \theta)$, and considered with respect to the colder portion, is similarly proportional to $m_2(\theta - \theta_2)$, Richmann's rule is accounted for or explained.

If we use q to represent a quantity of heat, and c to represent a factor of proportion, which may depend on the nature of the substance but which is as yet undefined, we may write

$$q = cm(\theta - \theta_0). \tag{92}$$

We have thus reached a conception of heat as something which may enter or leave a body, which is distributed throughout its mass, and which determines its temperature. We have further found a way to measure it, at least to measure so much of it as enters or leaves a body, by the observation of the mass of the body

and of its change of temperature. By selecting a particular body and a particular change of temperature, we may define a unit of heat.

243. The Unit of Heat. — A unit of heat which is frequently used in physical investigations is called the *calorie*. It is the heat which will raise the temperature of a kilogram of pure water one degree Centigrade. As recent observation has shown that the amount of heat which will raise the temperature of a kilogram of water one degree is slightly different in different parts of the scale, it is necessary, in order to give greater precision to this definition, to specify the particular degree on the scale through which the temperature of the water shall be raised. The degree usually chosen is that between 0° and 1° on the Centigrade scale, though other degrees have been chosen.

It is often convenient to use a smaller unit of heat than this, and we accordingly choose, as another unit, the heat which will raise the temperature of a gram of water from 0° to 1° Centigrade. This unit we may call the gram-degree, or the *therm*.

244. Heat Capacities. — When masses of two different substances, whose temperatures are different, are *mixed*, that is, are so brought together that they come to a common temperature by the passage of heat from one to the other, the resulting temperature does not conform to Richmann's rule. It was discovered by Black (1757) that the resulting temperature could be accounted for in a way consistent with our conception of heat, and expressed by a rule generally similar to Richmann's rule, if it is assumed that the effect of heat in raising the temperature of a body depends not only upon its mass but upon the substance of which it is composed. Black expressed his views by saying that each body has its own peculiar *capacity for heat*. By this he meant that the quantities of heat required to raise the temperatures of different bodies by the same amount depend not only on their masses but on characteristic factors, which are different for different substances. This is equivalent to saying that in the formula $q = cm(\theta - \theta_0)$, by which we have measured a quantity of heat (§ 242), the factor c is not a universal constant, but is characteristic or specific for each substance. The product cm for any body is the *heat capacity* of that body.

Using this conception of Black's and distinguishing the symbols referring to two substances by different subscripts, we may represent the quantity of heat transferred between the two bodies used

in the Taylor-Richmann experiment by the two equal expressions

$$c_1 m_1(\theta_1 - \theta) = c_2 m_2(\theta - \theta_2). \tag{93}$$

The test of the correctness of Black's views is obtained by mixing various different masses of any two chosen substances, whose initial temperatures are different in different trials, and by noticing that the masses and temperature changes are always so related that the ratio c_2/c_1 is a constant for any chosen pair of substances.

If the ratio of c_2 to c_1 is known, we may obtain the resulting temperature of such a mixture by a rule similar to Richmann's, in which the masses are replaced by their heat capacities. The formula

$$\theta = \frac{c_1 m_1 \theta_1 + c_2 m_2 \theta_2}{c_1 m_1 + c_2 m_2} \tag{94}$$

follows directly from the preceding one.

245. Specific Heats. — By mixing two bodies at different temperatures and observing the resulting temperature we may determine the ratio of the two characteristic factors c_1 and c_2 from the formula

$$\frac{c_2}{c_1} = \frac{m_1(\theta_1 - \theta)}{m_2(\theta - \theta_2)}.$$

If in all our experiments we use some one substance as standard, and determine the ratios of the characteristic factors of other substances to that of the standard substance, these ratios are characteristic of or specify the substances to which they pertain. They may be called the *specific heats* of the various substances referred to the standard substance.

In the determination of the specific heats of solids and liquids, the substance universally used as standard is water. If we arbitrarily assign the value unity to the characteristic constant for water, or set $c_1 = 1$, the characteristic constants of other substances have the same values as their ratios to the characteristic constant of water, so that they become the specific heats of the other substances.

This assignment of the value 1 to the specific heat of water is equivalent to adopting the unit of heat which has already been defined (§ 243). We have generally as an expression for the quantity of heat which enters a body, $q = cm(\theta - \theta_0)$; if the substance of the body is water and we set $c = 1$, then $q = 1$, or is the unit of heat, if the mass $m = 1$, and the rise of temperature $\theta - \theta_0 = 1$.

If we consider two quantities of heat q_2 and q_1, one of which enters a body of mass m having the specific heat c_2 and raises its temperature by $\theta - \theta_0$, and the other enters an equal mass of water and causes the same rise of temperature, and if we form the ratio between these quantities of heat we have

$$\frac{q_2}{q_1} = \frac{c_2 m(\theta - \theta_0)}{c_1 m(\theta - \theta_0)} = \frac{c_2}{c_1} = c_2,$$

since $c_1 = 1$. We thus obtain another way of defining the specific heat of a substance. As this formula shows, the specific heat of a substance is the ratio between the quantity of heat which will raise the temperature of a mass of the substance through any range to the quantity which will raise the temperature of an equal mass of water through the same range.

If the mass chosen is one gram and the rise of temperature chosen is one Centigrade degree, the denominator of the ratio is one gram-degree, or therm. The ratio, or the specific heat of the substance, is then otherwise defined as the quantity of heat, measured in therms, which will raise the temperature of a gram of the substance one Centigrade degree.

The *heat capacity* of a body is the product of its specific heat and its mass.

246. Heat Absorbed by Melting Solids. — In the investigation of specific heat it was found by DeLuc, and soon afterwards by Black, that the law by which the resulting temperature of a mixture is determined does not hold good if one of the bodies melts when it is brought in contact with the other. In any such case the resulting temperature is always lower than that given by the law. Black was therefore led to consider the process of melting in order to ascertain whether heat is required for it. In one of his experiments he placed two similar vessels, one containing water, the other containing an equal mass of ice, on the top of a stove. The temperature of the water began to rise at once. The ice in the other vessel melted, but its temperature, and that of the water which flowed from it, remained constant until melting was complete. Then the temperature in that vessel also began to rise, and rose at the same rate as in the other vessel. It is plain that heat must have been entering both vessels all the time and at the same rate, and since no evidence of its having entered the ice was given by any change of temperature, it must have been somehow used in melting the ice. In another of his experiments Black placed a

mass of ice in a mass of warm water, and determined the temperature which resulted when the ice was melted. His observations showed that the warm water lost more heat in coming to the common final temperature than was used in heating the water formed from the melting ice to that temperature, and he concluded that the difference was used in melting the ice.

The loss of heat in this case is especially apparent when the experiment is tried in a way indicated by Black. Equal masses of ice and water are taken, and the temperature of the water raised to 80° C. When the ice is immersed in the water, it begins to melt, and the temperature of the water falls more and more as the melting proceeds, until just as the last trace of ice disappears, the temperature of the mixture falls to 0° C. The warm water in this case has given up heat, and this heat has occasioned no rise of temperature in another body. We conclude that it has been used in melting the ice.

Black considered that when heat enters a body in such a way that it causes a rise of temperature, it is in such a condition with relation to the body that it can be detected by the temperature sense. He therefore called it sensible heat. On the other hand, the heat which has passed into ice, and melted it, cannot be detected by any change of temperature which it causes, and is in a sense concealed in the water. He therefore called it *latent heat.* This term is a very convenient one and is often used, although our present conception of heat as a form of energy makes it somewhat inappropriate. We may, instead of it, use the term *heat of fusion.*

247. Latent Heat or Heat of Fusion. — The absorption of heat by melting ice is an example of what occurs whenever a solid body melts. In every case of melting, a quantity of heat is absorbed which is proportional to the mass of the body melted. The factor of proportion is a characteristic constant for each substance. The observations which prove this are like the one described above, in which a mass of the solid at its melting temperature is immersed in a mass of the liquid at a higher temperature, and the resulting temperature observed. If we represent the masses of the solid and of the liquid by m_0 and m_1 respectively, the melting temperature by θ_0, the initial temperature of the liquid by θ_1, the final common temperature by θ, and the specific heat of the liquid by c, we obtain for the difference q between the heat which leaves the original liquid as it cools, and the heat which enters the liquid formed from the melting solid as it warms,

$$q = cm_1(\theta_1 - \theta) - cm_0(\theta - \theta_0).$$

Observation has shown that this difference q is proportional to

the mass m_0 of the solid which is melted. Introducing a factor of proportion L, we may write $q = Lm_0$, and using this in the last equation we have

$$cm_1(\theta_1 - \theta) = cm_0(\theta - \theta_0) + Lm_0. \quad (95)$$

The term on the left represents the heat which leaves the warm liquid; the first term on the right represents the heat which enters the cold liquid to warm it, the second term the heat used in melting the solid.

The factor L is called the *latent heat*, or the *heat of fusion*, of the substance to which it refers. It is found to have a characteristic value for each substance. It is evidently equal to the quantity of heat, measured in therms, which is absorbed by one gram of the substance as it melts. The latent heat, in ordinary circumstances, is large compared with the specific heat. Thus the latent heat of ice is about 80 therms.

248. Freezing Mixtures. — Black perceived that when a body melts, even though its melting is not brought about by the entrance of heat from without, it will of necessity absorb heat, and thus, if it does not receive heat from without, its own temperature will fall. He explained in this way the behavior of freezing mixtures, and showed that the general fact observed by Boyle, that all such mixtures melt, is the one upon which their efficacy as freezing mixtures depends.

249. Heat Emitted by Freezing Liquids. — Black also perceived that when a liquid freezes, it will give out an amount of heat equal to that which was absorbed by it when it was formed by melting. He explained in this way the constant temperature of a liquid while it is freezing, and also the rise of temperature which occurs when a supercooled liquid begins to freeze.

250. Heat Absorbed by Boiling Liquids. — Black also studied the case of boiling liquids. From the constancy of their boiling points he inferred that heat is absorbed by them during the process of boiling. The amount of heat, in therms, required to turn a gram of a liquid into vapor at the boiling temperature is called the *latent heat* of the vapor, or the *heat of vaporization* of the liquid. When the vapor condenses again, it may be made to heat a quantity of liquid of the same sort as that from which it is formed, and thus to give evidence that heat is emitted by a vapor on condensation. By taking advantage of the equality between the heat of vaporization and the heat emitted on condensation, the heat of vaporization may be determined. The heat of vaporization varies with the

temperature at which the evaporation takes place, being greater at lower temperatures than at higher.

251. Calorimeters. — A calorimeter is an instrument by means of which a quantity of heat may be measured.

From the principles developed in the previous sections, it is evident that we can measure the amount of heat which leaves a body by the effect which it will produce in some other body. It is not possible for us to measure all the heat which a body contains.

The calorimeter used for the *method of mixtures* consists of a vessel, isolated from surrounding bodies, so far as possible, so that the heat which is introduced into it will remain in it without change. A known quantity of water is placed in this vessel, and its temperature is taken. If another body of known mass is raised to a known temperature, higher than that of the water, and if it is then transferred to the water, its temperature will fall and the temperature of the water will rise, until they have reached a common value. This value is then determined. The quantity of heat lost by the body in falling from its original temperature to the common temperature is equal to the quantity of heat gained by the water in rising from its original temperature to the common temperature. The heat gained by the water is measured in therms by the product of the mass of the water and its change of temperature. Thus the heat lost by the body is determined.

When the method of mixtures is used, it is necessary to consider, in calculating the result of the observations, that the calorimeter containing the water has undergone the same temperature change as that of the water, and that some of the heat of the hot body has been employed in warming it. To take this into account it is usual to determine, by preliminary experiments or by calculation, the heat capacity of the calorimeter. This is a number which expresses the number of grams of water which has the same heat capacity, and as such is called the *water equivalent* of the calorimeter, or the *calorimeter constant*. It is added to the mass of water used to determine the mass heated by the hot body.

A calorimeter employed by Black and by Wilcke consists simply of a block of ice in which a small cavity is made. When the ice is at zero temperature, the interior of the cavity is dry. To keep it so, it is covered with a slab of ice. The body whose heat capacity is to be tested is heated to a known temperature and transferred to the cavity. The heat which it gives up will melt the ice around it. After its temperature has fallen to that of the ice, and

the ice no longer melts, the water which has been formed is taken out and weighed. We may state the result obtained in terms of an arbitrary unit of heat, namely, the quantity of heat required to melt a gram of ice. If this unit is used, the weight of the water obtained, in grams, measures the quantity of heat which the body has lost. As we know that 80 therms are required to melt a gram of ice, it is easy to state this quantity of heat in therms.

This *method of melting* has been applied in several different ways. The most ingenious of these is that invented by Bunsen, who utilized the change of volume, which occurs when ice melts, as a measure of the quantity of ice which is melted.

The *method of condensation*, which has been highly developed by Joly, measures the quantity of heat which will produce a given change of temperature in a body by weighing the amount of steam which is condensed upon that body.

The *method of cooling* measures the quantity of heat which leaves a body by the rate at which its temperature falls. In order to carry out the experiment on different substances under similar conditions, an instrument is constructed consisting of a small polished box, in the middle of which stands the bulb of a thermometer. The box is first filled with a standard substance. It is then raised to a high temperature and placed within a larger box, from which the air can be exhausted, and whose walls are kept at a constant temperature by immersion in ice. The thermometer is observed from minute to minute, and the rate at which the temperature changes for the standard substance is thus determined. Other substances are compared with the standard by carrying out similar observations with them.

The *method of comparison* depends upon the introduction of equal quantities of heat, in any way in which that may be done, into known masses of a standard substance and of the substance under examination and the observation of their changes of temperature.

252. Properties of Specific Heats. — Calorimetric observations are usually employed to determine either the heat capacity of a body, or the specific heat of a substance. The heat capacity of a body is supposed to be constant within the range of temperature employed in the experiment, and is therefore determined from the quantity of heat which leaves the body, when its temperature undergoes a known change, by dividing that quantity of heat by the change in temperature. The specific heat of the substance of

which the body is composed, provided it is homogeneous, is obtained by dividing the body's heat capacity by its mass.

By the study of the specific heats of solid substances through different ranges of temperature, it is found that, as a first approximation, they are constant at all ordinary temperatures. As a rule, the specific heat of a substance increases slightly as the temperature rises. There are a few substances, of which carbon, in the form of the diamond, is an example, whose specific heat increases rapidly with rise of temperature. The specific heat of the diamond is nearly three times as great at 200° as at 0°.

The specific heats of liquids also vary in a similar way with the temperature. The most reliable observations indicate a double variation in the case of water, its specific heat diminishing slightly from 0° to about 35°, and increasing from that point on.

The specific heat of a substance in the liquid state is always greater than that of the same substance in the solid state. The specific heat of water, which has been taken as standard, is greater than that of almost any other substance. So far as known, the only specific heats which are greater than that of water are those of hydrogen, and of mixtures of water with some of the alcohols.

The specific heats of a gas differ considerably according to the circumstances in which the measurement is made. If the gas is examined while its volume is kept constant, its specific heat will have a certain value, called its *specific heat at constant volume.* If, on the other hand, it is examined while its pressure is kept constant, so that as its temperature rises it expands, it is found that an additional quantity of heat is used in raising it to the same temperature, and its specific heat is greater than in the other case. The specific heat thus determined is called the *specific heat at constant pressure.* Regnault proved, by direct observation, that the specific heat at constant pressure is, in many cases, independent both of the pressure and of the temperature of the gas.

253. Dulong and Petit's Law. Atomic Heats. — A very remarkable relation among the specific heats of those chemical elements which are found in the solid state was discovered by Dulong and Petit (1819). These physicists, examining the specific heats of thirteen of the solid elements by the method of cooling, found that in each case the product of the specific heat by the atomic weight of the element was approximately the same number. Now the

masses of different elements which contain the same number of atoms are proportional to their atomic weights, so that the products of the specific heats and the atomic weights are the heat capacities of masses containing the same number of atoms, and since this product is the same for many solid elements, we conclude that for them their atoms all have the same capacity for heat. This law was subsequently shown to apply approximately to almost all the solid elements. The product here defined is called the *atomic heat* of the element. Its value is about 6.25. The atomic heats of a few of the solid elements, especially of carbon and silicium, are exceptions to the general rule.

It was shown by F. Neumann and by Regnault that the specific heats of substances which are compounds of the solid elements are such as to indicate that the atoms in composition retain their atomic heats. That is, it is found that the products of the specific heats and the molecular weights of compounds which have the same number of atoms in the molecule are approximately equal, so that the *molecular heats* of such compounds are equal. When the constituents of the molecules are elements which conform to Dulong and Petit's law, the quotient of the molecular heat divided by the number of atoms in the molecule is found to be the constant atomic heat already considered. When the molecule is one which contains atoms of an element which cannot be examined in the solid state, and other atoms of elements which conform to Dulong and Petit's law, we may calculate the atomic heat of the unknown element from the molecular heat of the compound. In this way the atomic heat of the gaseous elements, when they form parts of the molecules of solids, have been calculated. It is thus found that the elements hydrogen, nitrogen, and oxygen do not conform, at least in all cases, to Dulong and Petit's law. The organic compounds in general, which are compounds of carbon with these elements, have molecular heats which are not only inconsistent with the Dulong and Petit constant, but are inconsistent with any single constant values assigned for the atomic heats of the elements in the various compounds.

Transfer of Heat

254. Radiant Heat. — It is a matter of common observation that heat may be transferred from one body to another. Thus, when one end of an iron bar is thrust in the fire, the other end gradually gets warmer, and a body may be warmed by placing it in front of the

fire, although no part of it is in the fire. An experiment described by Newton proves that in the latter case the heat is not transferred from the fire to the body by the action of any known material body between them. Newton placed two thermometers in two similar glass vessels, from one of which the air was exhausted. After letting them stand in a cool place until the thermometers indicated the same temperature, he transferred them to a warm place, and found that the temperature of the thermometer in the vacuum rose nearly as fast as that of the other one, and that the final temperatures of both were the same. The heat which reached the thermometer in the vacuum was manifestly transferred to it from the walls of the vessel, when they were heated by standing in the warm place, and since there was no known material medium in contact with the bulb of the thermometer, the heat which it received must have been transferred to it without the intervention of any such medium.

Heat transferred in this way is said to be transferred by *radiation*, and is called *radiant heat*. Its properties are in every respect like those of light. Indeed, subsequent study has proved that radiant heat and light are essentially similar in all respects, and we shall therefore study it in connection with light. One general principle was discovered, however, governing the radiation of heat, which does not depend on the mode in which it is transmitted, and which may be considered in this place.

Taking advantage of the fact that radiant heat may be reflected and brought to a focus, as light is, the following experiment was tried: Two spherical mirrors were set up facing each other; a thermometer was placed at the focus of one of them, and at the focus of the other was placed a ball of lead which had been heated, though not to redness. When this was done, the temperature of the thermometer at the other focus began to rise, showing a reception of heat by the thermometer from the hot lead. When a block of ice was substituted for the lead, the temperature of the thermometer fell. Considering these experiments, Prevost perceived that the rational way to explain the fall of the thermometer produced by the ice was to ascribe it, not to a radiation of cold from the ice, but to a radiation of heat from the thermometer. Generalizing this idea, he laid down the principle that all bodies are at all times radiating heat, and receiving heat from neighboring bodies, and that the change of temperature of a body depends upon the relative amounts of heat which it is receiving and emitting. When it re-

ceives more than it emits, its temperature rises; when it emits more than it receives, its temperature falls; when it emits the same amount as it receives, its temperature is constant. This principle is known as *Prevost's law of exchanges.*

255. Conduction of Heat. — Heat which is transferred through a solid, when one part of it is heated, is said to be transferred by *conduction.* The experiments of Richmann and of Ingenhouss showed that the rate at which heat is transferred by conduction is different in different substances. It will manifestly depend also on the differences of temperature in the body, or on the way in which the temperature of the body changes along the lines along which conduction takes place. By assuming that the flow of heat along a line is proportional to the rate of change of temperature along that line, Biot, Rumford, and subsequently Fourier, were able to explain the movement of heat in bodies in a way which is consistent with the results of observation.

In accordance with the foregoing assumption, we may define the *conductivity* of a substance as the amount of heat which, in a unit of time, will pass between two unit areas in the substance, so placed that they stand at unit distance apart and that between them the temperature differs by one degree.

When heat passes across a surface at which two substances meet, its rate of transfer, or its surface conductivity, depends on the nature of the substances. As a first approximation, it is assumed to be proportional to the difference of temperature between the two substances. We may define the *surface conductivity* as the amount of heat which will pass, in a unit of time, through unit area of the surface, when the difference of the temperatures on the two sides of the surface is one degree.

In most cases in which bodies transmit heat by conduction, the temperatures of the different parts of the body will gradually approach definite values. After these definite values are attained, no more temperature changes occur. In all such cases it is plain that each part of the body receives as much heat from the hotter portions of the body as it sends on to the cooler portions. This condition of the body is called its steady state. In many other cases, in which the source of heat is not applied to the body continuously, the temperatures of its different parts vary continually in a way which depends on the way in which the heat is applied, on the shape and size of the body, and on its conductivities. Many such cases have been studied by the help of the assumptions already

described, and the theoretical results obtained have been found to agree with the results of observation.

256. Convection of Heat. — The transfer of heat in a liquid takes place generally by a process which is known as *convection*. It is a well-known fact that liquids receive heat readily. This was ascribed at first to liquids being very good conductors. Rumford noticed, however, that masses of liquid suspended in fibrous bodies take up heat very slowly, and retain it for a long time. This observation seemed to him inconsistent with the hypothesis that liquids are good conductors, and he accordingly undertook an investigation of the behavior of liquids when heated. Taking a glass flask with a long neck, he filled it with water, in which were suspended small particles or motes, and set it in a room whose temperature was low and constant. When it had stood there until its temperature had become that of the room, and until the particles suspended in the water were still, showing that there were no currents in the water, he transferred it to a warm room. The suspended particles at once began to move upward along the walls and to descend in the middle of the flask, showing that a regular circulation of the water was occurring. This circulation kept up until the temperature of the water had risen to that of the room. When he transferred the flask to the cold room again, currents in the opposite sense occurred, the water moving downward along the walls and upward in the middle. After these currents were once perceived, it was easy to explain them. When the flask was brought into the warm room, its walls took up heat from the surrounding air, and heated the layers of water which were near them. The density of the water was diminished by its expansion, due to this heating, and it therefore rose along the walls. The denser, because cooler, portions of the water which sank to give room at the top for the warmer portions, came in turn in contact with the walls, were also heated, and rose along the walls. In this way are explained the rapid reception of heat by a liquid, and the fact that the temperature of a liquid which is being heated from the bottom is almost the same throughout.

To test whether a liquid is a good conductor in the ordinary sense, Rumford tried to heat a mass of water by applying the heat at the top, so that the usual convection currents could not arise. He found so little heat transmitted to the bottom of the vessel that he concluded that water did not conduct at all. This conclusion is manifestly erroneous, for if there were no conduction in water,

the heating of the water near the walls of the flask, in Rumford's first experiment, could not be accounted for. As was shown by subsequent observation, water and other liquids are not nonconductors of heat, but very poor conductors.

Gases are ordinarily heated by convection. It was shown by Magnus that gases, like liquids, are poor conductors.

CHAPTER XIV

THERMAL RELATIONS OF BODIES

257. Expansion of Solids. — The general truth that a body expands when its temperature rises was illustrated in the early work on the subject of heat, by the expansion of the air in Galileo's thermometer, and by the behavior of thermometric substances in general. The Accademia del Cimento demonstrated that metals expand when they are heated. After the discovery of the general truth, it became a matter of interest to investigate the laws of this expansion. An instrument called the pyrometer was invented for that purpose, which consisted essentially of a framework so arranged that one end of the bar of metal under investigation could be kept fixed, while the other end was attached to a rack and pinion, by which a pointer could be moved over a dial. The rod was immersed in a vessel full of water or oil, which was heated from beneath. If the temperature of the water was raised and the bar expanded, the pointer moved around the dial, and thus measured the elongation of the bar. Inaccurate as this arrangement was, it served to show that the elongation of a bar of given length is proportional to the rise of temperature which occasions it, and that bars of different metals have different elongations for the same rise of temperature. This latter statement was illustrated by an experiment made by DeLuc, who clamped two bars, of iron and brass, firmly together at one end, and observed the relative expansion, when the bars were heated, by measuring the way in which the length of one bar increased more than the length of the other.

The first accurate observations of the absolute elongation of a bar were made by Laplace and Lavoisier. In their experiments, one end of the bar rested against a massive stone pier. The other end engaged with the short arm of a lever or system of levers, on the last long arm of which a mirror was mounted. While the bar was at the temperature of melting ice, the image of a vertical scale, reflected in this mirror, was observed by a telescope. The bar was then raised to the temperature of boiling water. The consequent elongation moved the levers, and so changed the position of the mirror. By another observation of the scale, and from a knowledge

of the ratios of the arms of the levers, it was possible to calculate the elongation of the bar. It was found that the elongation of unit length of the bar, produced by different changes of temperature, was not proportional to the change of temperature in each case. When the most accurate observations of this sort are made, it is found that to represent the length of the bar at different temperatures an expression involving the second, or even the third, power of the temperature has to be employed.

It is however true, as a first approximation, that the elongation is proportional to the rise of temperature. If we confine ourselves to this approximation, we may define the *coefficient of linear expansion* of a solid as the elongation of unit length of it when its temperature rises from 0° to 1° C.

Let l_0 and l represent the lengths of a bar at the temperatures 0 and θ respectively. If we represent the coefficient of linear expansion by α, we may represent the elongation of the bar, as its temperature rises from 0 to θ, by

$$l - l_0 = l_0 \alpha \theta.$$

The length of the bar at the temperature θ is then given by

$$l = l_0(1 + \alpha\theta). \tag{96}$$

The volume of a solid expands as its temperature rises in a way that depends upon its coefficient of linear expansion. Represent the three sides of a rectangular block at the temperature θ by a, b, c. Then in an easily understood notation the volume of the block will be given in terms of its volume at zero by the formula

$$abc = a_0 b_0 c_0 (1 + \alpha\theta)^3.$$

Now $(1 + \alpha\theta)^3 = 1 + 3\alpha\theta + 3\alpha^2\theta^2 + \alpha^3\theta^3$, and in all cases α is so small a quantity that the terms containing higher powers of it than the first may be neglected. This being so, and using v to represent volumes, we have

$$v = v_0 (1 + 3\alpha\theta). \tag{97}$$

The *coefficient of voluminal expansion* is thus three times the coefficient of linear expansion.

258. Expansion of Liquids. — When a liquid expands by heat in a thermometer bulb or in any similar vessel, the expansion which is observed by the rise of the column is the *relative expansion* of the liquid and the vessel. That is, the volume of the vessel increases as well as the volume of the liquid, and the change of volume indicated by the rise of the column is the difference between these two changes of volume. The *absolute expansion*, which it is sometimes important to determine, is the actual increase in volume of the liquid. DeLuc observed the relative expansions of various liquids and glass, and compared their absolute expansions by observing the rise of liquid columns in similar thermometer tubes. In the course of his observations he discovered a remarkable peculiarity in the

expansion of water. As the temperature of the water thermometer rose from the temperature of melting ice, the column in the tube at first fell, showing a contraction of the water. At 5° C., according to his observations, the column reached its lowest point, and from that temperature on it rose, showing a regular expansion. This observation has been repeatedly confirmed. The temperature at which the volume of the water is least, or at which its density is greatest, is really 4° C. This fact, together with the fact that ice is less dense than water, plays an important part in the economy of nature; for it is on that account that the temperature of the water in large ponds and lakes rarely falls below 4°, except near the top.

The most accurate study of the expansion of liquids has been made with the hydrostatic balance, by determining the apparent loss of weight of a standard body, like a hollow glass sphere, when immersed in the liquid at different temperatures. If the coefficient of expansion of the standard body is known, the coefficient of expansion of the liquid may be calculated from such observations. The *coefficient of expansion* in this case is defined as the increase in volume of unit volume of the liquid, when its temperature rises from 0° to 1° C. The volume calculated by this coefficient for any temperature will be only approximately correct. For accurate results, a more complicated temperature function must be used.

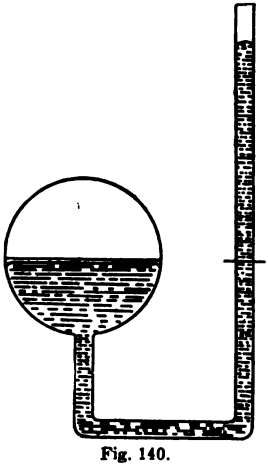


Fig. 140.

259. Expansion of Gases. Amontons' Experiment. — The first study of the expansion of gases, or of what amounts to the same thing, their increase in pressure with rise of temperature, was made by Amontons (1701). His instrument we may call an air-pressure thermometer (Fig. 140). It was a glass globe, into the bottom of which was inserted the short limb of a recurved tube. Mercury was introduced into the globe until it was about half filled, and so that the top of the column in the long limb of the tube stood at the same level as that of the mercury in the globe, when the instrument was at the standard temperature of melting ice. When this instrument was exposed to a higher temperature, the effect of the expansion of the enclosed air was to force down the mercury in the globe, and so to elevate the mercury column in the tube. Since the area of the mercury surface in the globe was very many

times greater than that of the cross section of the tube, the elevation of the mercury in the tube was as many times greater than the depression of the mercury in the globe. The air in the globe was thus subjected to pressure, proportional to the elevation of the mercury column, and its volume was maintained almost unchanged. If it had been kept exactly constant by a further increase of pressure, produced by adding more mercury to the column, the instrument would have been a perfect air-pressure thermometer. As it was, it served very well to enable Amontons to determine approximately the relation between the rise of temperature and the consequent increase in pressure. He stated that when the temperature rose from that of melting ice to that of boiling water, the increase in pressure was one-third the pressure at the lower temperature.

260. Expansion of Gases. Gay-Lussac's Law. — The later attempts which were made to obtain a measure of the expansion of gases were for a long time failures. It was shown by Dalton and by Gay-Lussac that these failures could be traced to the presence of water in the vessel containing the gas. By making the interior of the vessel perfectly dry, and by drying the gas, Gay-Lussac (1802) at last made a successful study of the expansion of gases. He stated his result as follows:

1. All gases, whatever be their density, and all vapors, expand equally for the same change of temperature.

2. For the permanent gases the increase from the ice point to the boiling point is 100/266.66 of the original volume.

The general law embodied in these statements is known as *Gay-Lussac's law*. The original factor given in Gay-Lussac's statement of the law has been slightly modified by subsequent observations. We now know that the increase in volume of a gas under constant pressure when its temperature rises from the melting point to the boiling point is 100/273 of the original volume. The coefficient of expansion of all gases is therefore 1/273. This statement is closely accurate for those gases which can be condensed only with great difficulty. Those which are easily condensed have generally higher values of the coefficient.

If we represent by v and v_0 the volumes of the gas at the temperature θ and 0 respectively, and use α as the coefficient of expansion, Gay-Lussac's law is expressed by the formula

$$v = v_0(1 + \alpha\theta), \tag{98}$$

with the unexpressed condition that the pressure remains constant during the expansion.

When the gas has expanded to the volume v, let the pressure be increased from its original value p_0 to p while the temperature is kept constantly equal to θ, and let p be such a pressure as to reduce the volume v to v_0. Then by Boyle's law (§ 180) we have $\frac{p}{p_0} = \frac{v}{v_0}$, and using this ratio in the formula above we obtain

$$p = p_0(1 + \alpha\theta). \tag{99}$$

This formula gives the law of increase of pressure with rise of temperature when the volume is kept constant. The coefficient of increase of pressure is the same as the coefficient of expansion. This conclusion is fully supported by the results of experiment.

261. Scale of the Ideal Gas Thermometer. — In much of our study of gases we consider a gas called the *ideal gas*, which has no precise counterpart in nature. One of the characteristics of an ideal gas is that it obeys Boyle's and Gay-Lussac's laws exactly.

We know from Boyle's law that the pressure of a gas is inversely proportional to its volume, and from Gay-Lussac's law that the pressure is directly proportional to the factor $1+\alpha\theta$, in which $\alpha = 1/273$ is the coefficient of expansion and θ is the temperature on the Centigrade scale. We may combine these proportions into an equation by introducing a factor of proportion, so as to have

$$p = c\frac{1+\alpha\theta}{v}.$$

If we substitute in this equation the numerical value of the coefficient of expansion, we obtain the equation

$$pv = \frac{c}{273}(273+\theta).$$

Now $c/273$ is a constant, to which we give the symbol R; and the quantity in parenthesis is the temperature on the Centigrade scale increased by 273. If, therefore, we suppose a thermometer so graduated that its zero indicates the temperature of $-273°$ C., and that the length of its degree is the same as that of the Centigrade degree, the temperature indicated by 0° C. will be indicated by 273 on the new scale, and any other temperature indicated by θ on the Centigrade scale will be indicated by $273+\theta$ on the new scale. We designate temperature given in the new scale by T. In terms of this new notation, we have the equation

$$pv = RT \tag{100}$$

as a statement of Boyle's and Gay-Lussac's laws.

262. Zero of the Ideal Gas Thermometer. — In the discussion of the kinetic theory of gases (§ 185) we obtained the equation $pv = \frac{1}{3}mnu^2$, showing that the pressure of a gas whose volume is

kept constant is proportional to the mean kinetic energy of its molecules. From the equation just obtained it therefore follows that the temperature of a gas, measured on this new scale, is proportional to the mean kinetic energy of its molecules. Now the least value which kinetic energy may have is zero, and it has no negative values. This relation therefore indicates that there can be no temperatures lower than the temperature indicated by the zero of the new scale, and that this zero is a limit of falling temperature. For reasons which will subsequently be given, it is called the absolute zero, and the scale, formed as we have described, is called the absolute scale of temperature. We have no warrant at present for the use of such terms. We shall accordingly call this zero the *zero of the ideal gas thermometer*, and the scale the scale of the ideal gas thermometer.

263. Air Thermometer. — A thermometer which will indicate this scale very nearly may be made by enclosing a mass of dry air in a vessel by means of a mercury column, so arranged that, as the volume of the air changes, the pressure on it may be kept constant. On this condition, changes in temperature will be proportional to the changes of volume. Such an arrangement, however, is not so satisfactory in its working as one in which the volume of the air is kept constant, as its temperature changes, by suitable changes of pressure. With this instrument the temperatures are taken proportional to the pressures, so that if the pressure on the air is determined for a known temperature, like that of melting ice, any other temperature may be obtained by an observation of the corresponding pressure.

264. Cooling of a Gas by Expansion. — A property of gases which is connected with the expansion with rise of temperature was discovered by Erasmus Darwin (1788) and investigated by Dalton (1802). They found that if air, or any other gas, which has been under pressure in a closed vessel, is allowed to expand suddenly into the atmosphere, it is cooled by the expansion, and that if it is suddenly compressed it is heated.

The heating produced by compression may be shown with the so-called fire syringe. This is a cylinder in which air is suddenly compressed by forcing down a closely fitting piston. The temperature produced by compression is high enough to cause the ignition of a bit of tinder placed in the cylinder.

265. Expansion of a Gas into Vacuum. — Another and more important general property of gases was investigated by Gay-Lussac,

and more satisfactorily by Joule (1845). Joule's experiment employed two stout metallic cylinders, in one of which air was compressed under high pressure. From the other the air was exhausted. The two cylinders thus prepared were connected by a tube, fitted with a stopcock, which could be opened to permit the compressed air to rush into the empty cylinder. They were immersed in water in a large calorimeter, and the temperature of the water was observed. The stopcock was then opened and the air was allowed to distribute itself uniformly in the two cylinders, when another observation of the temperature of the water showed no change in it.

To interpret this experiment we must anticipate the conclusion to which we shall come later, that heat is a form of energy. Using this principle we may argue that, since the temperature of the water did not change, no energy in the form of heat entered or left the air, and furthermore, as the air passed into the empty cylinder without doing external work, it did not acquire or lose mechanical energy. We conclude that its total energy remained the same in the larger volume as it had been in the smaller one, and generally that the energy of a gas does not depend on its volume.

Further, since the temperature of the air did not change, we know from Boyle's law that the product of its pressure and volume did not change. We also know from the kinetic theory of gases (§ 185) that this product is proportional to the mean kinetic energy of the molecules of the air. The mean kinetic energy of the molecules of the air therefore did not change, and comparing this with the constant energy of the air measured directly, we infer that the energy of the air consists in the kinetic energy of its molecules. We may generalize this conclusion for all gases.

Now the kinetic energy of the molecules of a gas changes when the temperature changes. We may therefore state as the result of Joule's experiments that the energy of an ideal gas is a function of its temperature, but is independent of its volume.

Later experiments by Joule and Lord Kelvin showed that a slight fall of temperature usually occurs when a gas expands without doing external work. The law stated holds for the ideal gas.

The cooling which occurs in these cases is ascribed partly to work done against the molecular forces between the molecules of the gas. The ideal gas is one in which the molecules do not exert forces on one another in any finite part of their paths.

266. Evaporation. — If water is exposed in an open vessel to the air it will gradually disappear. It is said to have evaporated, and

the process by which it disappears is called *evaporation*. Many other liquids evaporate as readily as water does, or even more readily. On the other hand, there are many liquids, like sulphuric acid or mercury, which evaporate so slowly that their evaporation can hardly be detected. If the temperature of water is observed while it is evaporating, it will be found to be lower than that of the surrounding air. If the water is placed in the receiver of an air pump, and the air around it exhausted, evaporation will go on much more rapidly than under atmospheric pressure, and the difference of temperature between the water and surrounding bodies will be much greater. During the process heat is entering the water from without, and the rate at which it enters depends on the difference of temperature between the water and surrounding bodies. We therefore infer that the process of evaporation involves the absorption of heat, and that the amount of water evaporated at a given temperature is proportional to the amount of heat which enters it. This conclusion has been confirmed by many observations.

267. Vapors. — The result of the evaporation of water is the production of another body, which is called the *vapor* of water, or water vapor. In general it seems to be similar to a gas. We may study the properties of a vapor by inserting a small quantity of the liquid, whose vapor we wish to examine, under the open end of a filled barometer tube (Fig. 141). The liquid will rise through the mercury of the column, and will evaporate in the vacuum above. Ordinarily, when the space above is not too great, some of the liquid will remain unevaporated and will float on top of the mercury column. As soon as the vapor is formed, the mercury column is depressed. This indicates that the vapor is exerting a pressure upon the top of the column. The change in height of the mercury column is a measure of this pressure. If the tube stands in so deep a vessel, and is itself so long, that we may raise or lower it so as to make considerable changes in the volume of the space above the column, we find, on changing the volume, that the pressure indicated by the height of the column above the mercury surface outside the tube remains constant. This is not the way in which a gas would behave, for its pressure changes when its volume

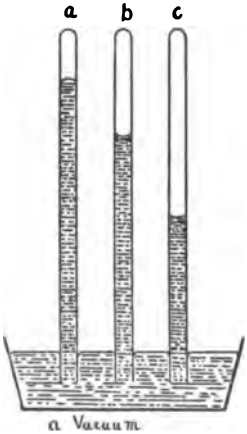


Fig. 141.

changes. If so little liquid has been introduced into the tube that none of it remains unevaporated, the pressure will be less than that obtained when liquid is present, and it will not remain constant when the volume is changed. The vapor in this case will behave like a gas. It is only when liquid as well as vapor is present in the tube that the constant and maximum value of the vapor pressure is exhibited. This maximum pressure is called the *vapor pressure* of the substance.

If the instrument, arranged as described, is exposed to a higher temperature, the top of the mercury column will be depressed, showing an increase in the vapor pressure. On the other hand, a cooling of the instrument will show a decrease in the vapor pressure. The rate of increase or decrease is different for different liquids, and has not been found to be expressible in any simple law.

268. Saturated Vapors. — A vapor in the condition in which it exhibits its maximum pressure is called a *saturated vapor*. We may restate the experimental results already described by saying that the pressure of a saturated vapor is a function of its temperature, and is independent of the volume occupied by the vapor. We may explain this law by supposing that a saturated vapor is one in which every unit of volume contains as much vapor as can exist in it at the given temperature without condensation. If the volume of the vapor is diminished, enough of the vapor condenses into the liquid state to keep the density of the remaining vapor the same as before. If the volume is increased, enough of the liquid evaporates to saturate the larger volume.

When the volume of the vapor is kept fixed, while its temperature is raised, the increase in pressure exhibited by it is not due simply to the increased energy of its molecules, but also to the fact that more of the liquid evaporates, so that the density of the vapor is increased. On the other hand, when the temperature is lowered, some of the vapor is condensed.

When one part of the surface of any vessel containing a saturated vapor is cooled, condensation takes place on it and the pressure near it is lowered to the saturation pressure corresponding to the lower temperature. The vapor in the rest of the vessel streams toward the place of low pressure and is also condensed, so that the pressure throughout the vessel becomes that corresponding to the lowest temperature of any part of it. The process of distillation takes advantage of this fact.

269. Dalton's Law. — It was discovered by Dalton that when vapors of two different liquids are formed in the same volume,

the maximum pressure which they exert is equal to the sum of the maximum pressures exerted by the two vapors separately. The same thing is true for mixtures of several vapors, or, if the volume is kept constant, for mixtures of gases, or of vapors with gases. In all these cases the pressure of the mixture is equal to the sum of the pressures of its constituents, if they were to occupy the same volume separately. This law is known as Dalton's law.

270. Ebullition. — The evaporation which has been described takes place at the free surface of the liquid. Another mode of evaporation occurs, called *ebullition* or *boiling*, in which the evaporation occurs within the body of the liquid. When the temperature of the liquid has reached the point at which its vapor pressure equals the external pressure, bubbles of vapor appear, arising at points on the wall of the containing vessel. After boiling has fairly begun, these bubbles rise through the liquid, rapidly increasing in volume as they do so, and break at the top, liberating the vapor which they contain. The temperature at which this process takes place generally depends, for a given liquid, on the pressure upon it, but it has been shown, by Dufour, that if the liquid is prepared by previous boiling so that the air ordinarily dissolved in it is driven out, its temperature may be raised considerably above its ordinary boiling point, without its boiling. When boiling begins in this case, the formation of vapor is exceedingly rapid. These phenomena are analogous to those observed when a liquid is supercooled.

271. Kinetic Theory of Evaporation. — The striking similarity of vapors and gases suggests that it may be possible to explain the properties of vapors by the kinetic theory. In order to do so, we extend the kinetic hypothesis to liquids, so far as to assume that the molecules of a liquid are in motion, and that they are not all moving with the same velocity. Their average velocity will have a value depending on the temperature, but the velocities of the separate molecules may, in some cases, very much exceed this average velocity. We apply this hypothesis to a liquid exposed in a closed and otherwise empty vessel. Among the molecules of the liquid, which are moving in various directions, there will be some near the upper surface which are moving upward, and some of these will be moving with velocities which are high enough to carry them beyond the attraction of the neighboring molecules of the liquid. They then enter the space above as molecules of vapor.

Now the molecules of the vapor are also moving in various

directions, and some of them will come within the range of the forces of the liquid, and will return to it. The vapor will attain its greatest density when it is so dense that the molecules which enter the liquid from the vapor are equal in number to those which enter the vapor from the liquid. The attainment of this condition plainly does not depend on the total number of molecules in the vapor, but only on the density of the vapor just above the liquid, and it therefore follows that the maximum pressure of the vapor is independent of its volume. Since the velocity of the molecules increases as the temperature rises, there will be more molecules at the higher temperature whose velocities will be sufficient to carry them away from the liquid, and consequently the density of the vapor above the liquid will have to be greater, at the higher temperature, in order that as many molecules may leave the vapor as enter it. Therefore the density and the pressure of a saturated vapor will increase on rise of temperature.

272. Critical Temperature. — Caignard de la Tour tried the experiment of heating ether when it was sealed up in a strong glass tube. Under these conditions, as the temperature rose, the vapor in the tube became more dense. The surface of separation between the liquid below and the vapor above, called generally the meniscus, remained distinctly visible until a certain temperature was reached. At that temperature it disappeared, and the contents of the tube became apparently homogeneous. Before the disappearance of the meniscus, its position in the tube indicated the presence of a considerable mass of liquid, and its disappearance did not seem to be due to gradual evaporation, but to the attainment of a condition in which the liquid did not differ in appearance from the vapor above it. At higher temperatures the contents of the tube seemed homogeneous. The temperature at which this change occurs is called the *critical temperature.*

When the tube was cooled again, and the critical temperature was reached, a sudden condensation of a fine fog or rain took place throughout the tube, and the liquid reappeared.

Similar experiments were tried with other liquids, and for many of them the existence of a critical temperature was determined.

273. Liquefaction of Gases. — If the pressure on a nonsaturated vapor is gradually increased, the vapor will at last reach the condition of saturation, and liquefaction will begin. This fact suggested the possibility of liquefying the gases by increasing the pressure upon them.

For some time attempts made in this direction were unsuccessful. Although the gases in some cases were subjected to enormous pressures, Natterer, for example, using pressures as high as three thousand atmospheres, they showed no signs of liquefaction. Among the first successful experiments were those made by Faraday (1823). The gases to be examined were generated in a strong glass tube, bent in the middle. Substances which would produce the gas by chemical action were introduced in one of the limbs of this tube, and the tube was then sealed. As the gas was continually generated, the pressure on it increased. Taking a hint from de la Tour's observations, Faraday surrounded the end of the tube which did not contain the substances generating the gas with a freezing mixture, in order, if possible, to lower the temperature of that part of the tube below the critical temperature. With this arrangement he liquefied chlorine, carbon dioxide, and many other gases. A few of the gases, among them hydrogen, nitrogen and oxygen, proved refractory to this treatment. Faraday recognized that the reason of this was his inability to obtain a temperature as low as the critical temperatures of these gases.

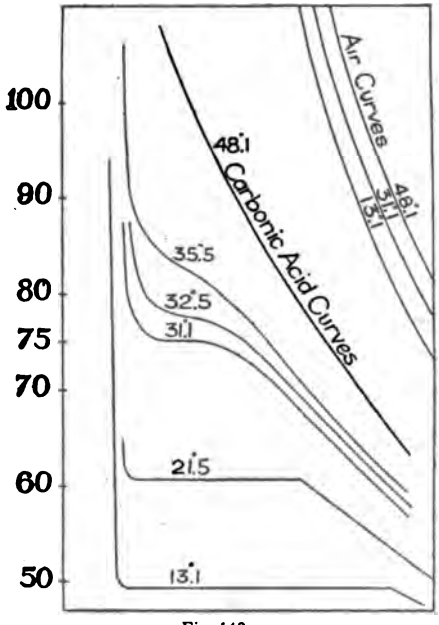


Fig. 142a.

274. Conditions Requisite for Liquefaction. — Andrews (1863, 1869) very much clarified the views which were entertained as to the conditions requisite for the liquefaction of gases by his study of the behavior of carbon dioxide under pressure. By confining the gas in stout capillary tubes, he was able to subject it to high pressures, and to measure the corresponding volumes at different constant temperatures. The results of these measurements are given in a series of diagrams or curves, plotted with reference to two rectangular axes. Volumes are measured along the axis of abscissas, pressures along the axis of ordinates. Each curve represents the relations of volume and pressure

as these vary while the temperature is kept constant. Because of this condition the curves are called *isothermals*.

When Andrews' diagrams are examined, it appears that the isothermals for temperatures below 30.9° C. are different in form from those for temperatures above that limit (Fig. 142a).

To describe Andrews' results more clearly it is advisable to refer to a more complete diagram (Fig. 142b), in which the isothermals are represented in a form which, while not exactly that of Andrews' curves, is yet perfectly typical. Let us consider the curve marked by letters in the diagram. When in the condition represented by the point *a* on that curve, the carbon dioxide is to all appearance a gas. As the pressure is increased and the volume correspondingly diminished, the successive corresponding values are given by the points of the curve lying between *a* and *b*. The curve between these points slopes upward to the left and indicates an increasing pressure with diminishing volume, generally according to the behavior of the more refractory gases. After the condition indicated by the point *b* is attained, further diminution of volume takes place without increase in pressure, and the substance in the tube is evidently present in both the liquid and the gaseous states. The gas in this condition is simply a saturated vapor. By a diminution of the volume at constant pressure to the volume indicated by the point *c* the gas is entirely liquefied. From that point on any further diminution of the volume necessitates a great increase in pressure. The curve from the point *c* on slopes upward to the left very rapidly, and indicates a rapidly increasing pressure with diminishing volume, corresponding to the behavior of other liquids.

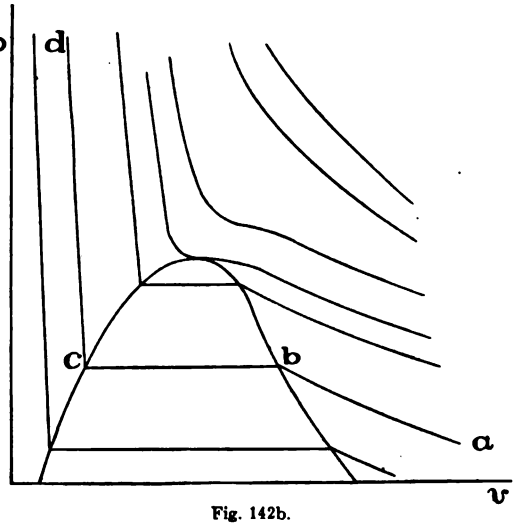


Fig. 142b.

Other isothermals drawn for temperatures below 30.9° exhibit similar features. They agree in showing a part on the right which

corresponds to the gaseous state, a horizontal part which corresponds to the state of saturated vapor and its liquid, and a part on the left which corresponds to the liquid state. When the whole set of isothermals is examined, it will be noticed that as the temperature to which the isothermal corresponds is higher, the two terminal points of the horizontal part are nearer together. This indicates that the volume of the gas as a saturated vapor and the volume of the gas when liquefied approach each other as the temperature rises. A line is drawn in the diagram through the terminal points of the horizontal parts of the various isothermals corresponding to temperatures at which liquefaction can take place.

The isothermal which touches the summit of the curve indicated by that line corresponds to the critical temperature 30.9°. The pressure and volume indicated by the point of tangency are the critical pressure and volume. The gas when in the condition indicated by that point seems to waver in its properties between those of a liquid and those of a gas. If its temperature is in the slightest degree raised it is unmistakably a gas. If lowered, a large part of it is unmistakably a liquid. As the temperature is raised past the critical temperature the liquid suddenly, and not apparently by ordinary evaporation, becomes similar to the vapor above it, and the contents of the tube seem homogeneous and gaseous.

The isothermals corresponding to temperatures above the critical temperature show no horizontal part. Diminishing volume at these higher temperatures is always accompanied by an increase of pressure. In the neighborhood of the critical point the curves show a peculiar shape indicated in the diagram, but for the most part they are similar in general to those drawn for the more refractory gases. At very high temperatures all such peculiarities disappear from the curves.

These experiments of Andrews indicate very clearly the existence of a critical temperature marking the highest temperature at which a gas or vapor can be condensed into a liquid, and marking a limit above which a gas or vapor cannot be condensed into a liquid by any pressure whatever.

An explanation of the phenomena at the critical temperature is sometimes given by supposing that that temperature marks the point at which the liquid and its vapor, while not losing their distinctive properties, become mutually soluble in each other.

275. The Continuity of the Liquid and Gaseous States. — Andrews interpreted his experiments as showing that there is an

essential continuity between the liquid and the gaseous states of matter. In the experimental examination of the transition from liquid to vapor there appears to be an abrupt change in the shape of the isothermal at the point where vaporization begins, and again at the point where vaporization is complete, but this is interpreted as a sort of average result due to changes in elementary volumes of the substance which are really continuous.

James Thomson represented the transition from the liquid to the gaseous state by an isothermal curve of the sort shown in Fig. 143. The ordinate of the horizontal line p' represents the pressure at which the liquid will exist in equilibrium with its vapor at the temperature of the chosen isothermal. The abscissas of the points f and b represent the volumes of the mass chosen when vaporization is just beginning and when it is complete. The portion fe of the curve represents a condition of the substance in which it is liquid under a lower pressure than the normal vapor pressure for the given temperature. The portion cb represents a condition of the substance in which it is a vapor or gas under a higher pressure than the normal vapor pressure. The portion ec, in which the pressure and volume both increase together, represents an unstable state of the substance, in which it is in transit from the liquid to the gaseous state.

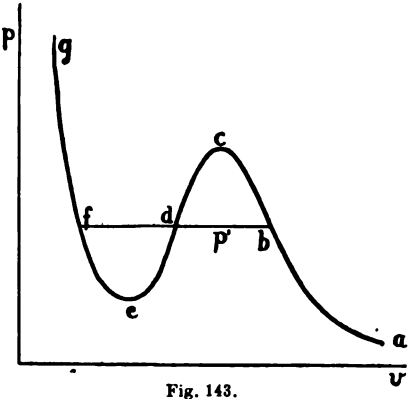


Fig. 143.

The existence of a condition corresponding to that represented by points on the line fe was demonstrated by Dufour. After boiling water in a vessel until the air had been expelled from it, and allowing it to come to rest at a certain temperature, Dufour found that it was possible to remove some of the air above the water by an air pump, and so to reduce the pressure upon it, until the pressure was less than that at which water would normally boil at that temperature, without the water's boiling.

Dufour also suspended small globules of water in a mixture of oils, and found that he could then raise the temperature far above the boiling point corresponding to the pressure upon the surface without the water's boiling.

The condition corresponding to the portion *cb* of the curve has also been experimentally demonstrated by showing that it is possible to obtain a supersaturated vapor by freeing the space containing it of dust and other nuclei upon which condensation will otherwise begin.

Van der Waals has proposed a formula to represent the true isothermals of a liquid and its vapor, which is a modification of the Boyle-Gay-Lussac formula. He conceived of the molecules of gas or vapor as exerting attractive forces on one another, and therefore virtually diminishing the volume of the gas in the way in which an additional external pressure would do. He showed that this additional pressure would be inversely as the square of the volume of the gas, and so added to the pressure in the simple formula the term a/v^2, in which a is a constant which is characteristic of the gas or vapor. He further showed that if the molecules have appreciable volume, their free paths will be less than those calculated on the supposition that the molecules are negligibly small, and that the effect of this is equivalent to having the molecules negligibly small and moving about in a smaller volume. He therefore diminished the volume factor in the simple formula by a constant b.

The van der Waals formula has the form

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT. \tag{101}$$

The constants a and b can be determined by a study of the observations of the isothermals of a gas or vapor. The formula gives a curve for the isothermal which is of the same general form as the hypothetical curve of James Thomson. By finding the value of T when the three roots of this cubic equation in v are equal, the critical temperature may be calculated in terms of the constants for the gas. A very good agreement with observation is found in many cases in which the van der Waals formula has been applied.

276. Liquefaction of Gases. — The refractory gases, or at least some of them, were condensed by Cailletet and by Pictet (1878). The methods used by these investigators were essentially the same. The gas to be condensed was forced under high pressure into a small chamber furnished with a stopcock opening to the air. After the gas had been cooled by refrigerating agents to the lowest attainable temperature, the stopcock was opened, and it was allowed to rush out. The additional cooling, produced by the sudden expansion, brought the temperature down below the critical

value, and a little of the gas condensed as a liquid, under atmospheric pressure, on the walls of the tube. In this way nitrogen, oxygen, and other gases were first liquefied.

By using more powerful refrigerating agents, Olszewski, Dewar, and others have succeeded in lowering the temperature below the critical value, while the gas is under pressure, and in this way have obtained large quantities of nitrogen, oxygen, and other gases in the liquid state. By allowing the liquid thus obtained to evaporate rapidly, its temperature is still further lowered. In this way nitrogen has been frozen at $-220°$ C.

Kamerlingh Onnes succeeded in liquefying helium by a method which is essentially the same. This gas, of atomic weight 4, is of peculiarly simple structure, has but one atom in the molecule, and seems not to enter into chemical combination with atoms of other elements. It is apparently identical with the so-called α-particles emitted from the radioactive elements, after they have lost the electric charges with which they are connected when they leave the radioactive element. By using large quantities of liquid hydrogen to cool the helium under pressure to a low temperature and by then relieving the pressure on the gas, it was cooled below its critical temperature and was liquefied. The liquid helium obtained boiled under atmospheric pressure at $-268°$ C., and by rapidly removing the gas as it was formed by evaporation, the temperature of the remaining liquid was lowered to $-271.4°$ C.

By taking advantage of the cooling of a gas produced by its steady flow from an orifice without doing external work (§ 265), and by cooling the gas which is still under pressure by means of the cooled gas carried back past the pressure chamber, Linde has succeeded in liquefying air in large quantities under atmospheric pressure. The same method has also been used for the liquefaction of other gases.

CHAPTER XV

HEAT AND ENERGY

277. Heat a Mode of Motion. — In what we have studied up to this time we have ignored the question of the nature of heat. In describing the phenomena and in stating the laws of the phenomena, it has not been necessary for us to know more about heat than that it is something which is associated with the sensation of temperature, and which can be measured by the calorimeter. In this respect our course has been similar to that which was followed in the actual development of the subject, for although the conception of heat which prevailed during the time of that development was given greater precision of statement than we have found necessary, yet after all, it amounted to nothing more than that which we have used. During the century in which most of the experimental work was done, heat was commonly believed to be an indestructible material or substance, whose passage from one body into another was generally indicated by changes of temperature or of physical state. This substance was supposed to be divided into minute particles, which repelled one another and attracted the particles of ordinary matter. Since bodies do not change in weight when heated, it was supposed not to be subject to the force of gravity, and was called imponderable. The substance thus conceived of was named caloric by Laplace and Lavoisier. It is easy to see that by ascribing the right sort of properties to such a substance, most of the phenomena which we have described can be explained.

A few facts, however, were known, of fundamental importance for the theory, which could not be explained on this hypothesis. Of these the most important and general one is the production of heat by friction. Lord Bacon called attention to this, and made it one of the strongest points of the argument in which he contended that heat was not a substance, but the motion of the minute parts of bodies.

In 1798 Count Rumford published an account of experiments which he had made on the heat produced by friction. He was at the time engaged in superintending the making of cannon in the

arsenal at Munich. His attention was attracted to the great quantity of heat developed in the cannon when it was being bored out, and especially by the high temperature of the chips cut out by the boring tool. The upholders of the caloric theory of heat explained this high temperature by supposing that the specific heat of the metal was diminished by the friction upon it, so that the same quantity of heat in it would cause a higher temperature. Rumford first showed by direct measurement that the specific heat of the chips cut off by the tool was the same as that of any block of the same metal. He then mounted a cylinder of the metal in such a way that it could be revolved around a blunted boring tool, which pressed against the bottom of a deep cavity cut in it. With this arrangement he found that as the block of metal was turned its temperature continually rose. The amount of metal abraded by the blunt tool was so trifling that the rise of temperature could not reasonably be ascribed to any change in its specific heat, and Rumford concluded that it was due to heat directly produced or made by the friction. By surrounding the block of metal with water, he was able to measure the quantity of heat thus produced. In Rumford's mind the essential feature of the experiment was the continued production of heat in apparently unlimited quantities, without any change occurring in the bodies producing it by which it could be accounted for. He stated his conclusion in the following words: "It is hardly necessary to say that anything which any insulated body, or system of bodies, can continue to furnish without limitation, cannot possibly be a material substance; and it appears to me to be extremely difficult, if not quite impossible, to form any distinct idea of anything capable of being excited and communicated in the manner the heat was excited and communicated in these experiments, except it be motion."

Almost at the same time Davy performed experiments of a similar nature. He constructed a mechanism by which two blocks of ice were rubbed together in a room whose temperature was below the melting point, and found that by this operation the ice was melted. He also constructed an arrangement by which a brass disk was turned by clockwork against the friction of another piece of brass. Small pieces of wax were placed on the disk, to show, by their melting, when the temperature had risen to the melting point of the wax. This apparatus he placed on a block of ice in the receiver of an air pump, and exhausted the air from around it. He found that, when the disk was turned for a while, the wax was

melted, and the arrangement of the experiment was such that the heat which appeared could not have entered the disk from without. He drew from these observations essentially the same conclusion as that drawn by Rumford.

These experiments of Rumford and Davy were accepted by some as proving that heat could not be a substance, and therefore that, as Thomas Young said, "it must be a quality, and this quality can only be motion." In general, however, they had no effect in changing the prevailing theory that heat was a substance. The holders of that theory either ignored them altogether, or set them aside with the expectation that in the future they would be explained in a way consistent with their theory.

278. Heat a Form of Energy. — In 1842 Robert Mayer published a paper in which he asserted the possibility of the transformation of mechanical energy into heat, and the reversed transformation of heat into mechanical energy. His argument was rather a metaphysical than a physical one, and attracted no attention. He took, however, one important step by attempting the calculation of the amount of heat which is equivalent to, or which may be transformed into, a unit of mechanical energy.

In the following year, Joule began an investigation of the various ways in which heat may be produced by the expenditure of mechanical energy, in order to determine whether or not the same amount of heat, in whatever particular way it may be produced, is always produced by the expenditure of the same amount of mechanical energy. The results obtained by the different methods which he employed were not very consistent with one another, though they were of the same order of magnitude. They were sufficiently consistent, however, to convince Joule that heat may be produced by the expenditure of mechanical energy, and that the amount of heat produced is always in the same proportion to the amount of energy expended. From the equivalence between heat and mechanical energy, he concluded that heat itself is a form of energy.

279. Mechanical Equivalent of Heat. — The ratio of the energy expended to the amount of heat produced by it, or the amount of energy which will produce one unit of heat, is called the *mechanical equivalent of heat.* Joule undertook the task of determining this quantity. To do this he used a vessel filled with water, in which a paddle wheel could be revolved (Fig. 144). Flanges projecting from the walls of the vessel, between which the blades of the paddle passed, broke up the circulation of the water, and greatly increased

the friction against which the paddle turned. The paddle was kept in motion by a falling weight. The mechanical energy expended, or the work done by the falling weight, was determined from the known value of the weight and of the distance through which it moved, and the heat developed by the friction of the paddle in the water was determined by an observation of the rise of temperature. It was found that whatever work was done, the ratio of the work done to the heat developed was constant. In order to vary the conditions of the experiment, Joule used a similar apparatus in which the vessel and paddle were made of iron, and the liquid was mercury, and also an apparatus in which the friction which developed heat was that between two iron plates rubbed together under mercury. From these three forms of the experiment he obtained the result that the amount of work which will raise the temperature of a pound of water one degree Fahrenheit is 772 foot pounds. This statement is equivalent to the statement that the amount of work which will produce a calorie of heat is 423 kilogram-meters. More recent determinations have increased these numbers somewhat. For ordinary work we take 427 kilogram-meters as the mechanical equivalent of one calorie. In absolute units 42×10^6 ergs is equivalent to one therm.

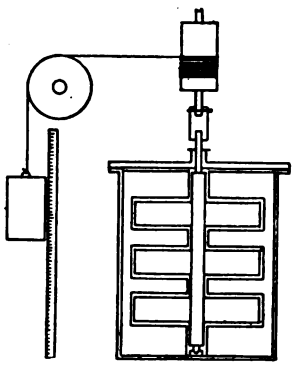


Fig. 144.

Joule's results carried conviction that heat is a form of energy, and that it can be transformed into mechanical energy, or mechanical energy transformed into it, without loss.

280. Conservation of Energy. — Heat may be transformed into work, or work into heat, by processes which are far more complicated than the one which we have described. In certain stages of these processes the condition of the system is often such that we do not recognize in it the presence either of mechanical energy or of heat. Whatever this condition may be, it is found that a quantitative relation exists between it and the work used, or the heat absorbed, to bring it about.

The first of Joule's experiments is a good example of one of these processes. Joule began by investigating the relations of heat to the electric current, and to the chemical action which takes place in the voltaic cell while the current is flowing. The chemical effect in the cells which he used was the consumption of

zinc by the acid of the cell. He first determined the amount of heat developed during the consumption of a certain mass of zinc by direct chemical action. He then determined the amount of heat developed in an electric circuit when the same amount of zinc was consumed. On comparing them he found that the two quantities of heat were equal. The difference between them was simply one of distribution, the heat in the first case being developed directly at the place where the chemical action was going on, while in the second case it was distributed throughout the circuit. Joule next inserted in the circuit a small magnetic motor. When the motor was at rest, or when it was revolving without doing any work, the heat developed in the circuit when the same amount of zinc was consumed was the same as before. When, however, the motor was made to lift a weight and so to do work, less heat was developed in the circuit. The heat which did not appear in this case had evidently been transformed into mechanical work. Roughly speaking, the ratio between the mechanical work done and the heat which had apparently been transformed into it was the same as the ratio found in Joule's later experiments, in which work was transformed into heat.

The process by which this transformation took place was not, however, so direct a one as that in which heat is produced by friction. It involved a series of intermediate conditions of the system, which were neither mechanical nor thermal conditions.

To explain this series of conditions, and countless others like them, we assume that energy exists in the universe in several forms, and that the energy in any one form can be transformed into any other. Thus, in the example before us, the energy of possible chemical combination, resident in the zinc and the acid, was transformed into electric energy, and this, in turn, either into heat or into mechanical work. So far as we can judge by experiment, transformations of energy take place without any loss or gain of energy.

The general principle that there exists in the universe a certain quantity of energy, which is unchanged by any natural operation, was first announced by Mayer. It was also announced by Joule on independent grounds, and received its first confirmation from his experimental work. The publication by Helmholtz, in 1847, of an important paper, in which he discussed the principle in connection with its application to various departments of physics, and showed that it was consistent with all that was known about natural operations, may be considered to mark the establishment of the principle.

In its general form, the principle is known as *the conservation of energy*. It may be stated as follows: In any closed system, that is, in any system into which no energy enters and out of which no energy goes, the amount of energy remains constant, whatever be the transformations going on within the system. To make this

statement accord more nearly with the conditions under which experiments must be carried on, we may say that the amount of energy in any system, whatever be the transformations which go on within the system, is increased or diminished only by an amount equal to that of the energy which passes through the bounding surface of the system. A statement which we cannot contradict, although it is one which we cannot prove, is that the energy of the universe is a constant quantity.

281. Kinetic Theory of Heat. — As an example of the use of the inductive method of reasoning, Lord Bacon, in the *Novum Organum*, considered the nature of heat. On grounds which would now be considered very unsatisfactory, he came to the conclusion that heat was a peculiar kind of motion of the small particles of bodies. Newton, and his contemporary, Locke, held the same view. The theory that heat is an imponderable substance was, however, so easy to work with, and satisfied so fully most of the demands made upon it by the discoveries of the time, that it displaced the other theory, except in the minds of a few leading thinkers, like Cavendish and Thomas Young. We have seen how Rumford and Davy were led to adopt a view similar to Bacon's by reflecting upon the results of their experiments. Joule's thought on the matter took the same course. When it was proved by Clausius and Maxwell that the properties of gases could be explained by the kinetic theory, it became almost a matter of course to extend this theory to all the states of matter, and to account for their relations to heat by the motions of their molecules.

The method employed by Mohr may be used to confirm our belief in the general kinetic theory of heat. It consists in assuming that the molecules of all bodies are in motion, and that this motion is increased when heat enters the body, and in tracing the consequences of this assumption to see whether it will afford reasonable explanations of the properties of bodies with respect to heat. For example, we may explain the expansion of bodies by heat by ascribing it to their increased molecular motions, by means of which each molecule is able to force its neighbors away from itself a little farther. We explain conduction of heat as the transfer of kinetic energy by impact from the hotter or more rapidly moving molecules to the colder ones near them. We explain evaporation by supposing that the molecules of the liquid leap away from its surface and escape from the attractions of the molecules remaining in the liquid. We explain the cooling of a gas when it expands

against external pressure, by supposing that the work that is done by that expansion is derived from the kinetic energy of its molecules. We explain the sensation of heat by supposing it due to the impacts of the molecules against the ends of the nerves. In general, when mechanical work is transformed into heat, we suppose it to be transformed into the kinetic energy of the molecules of the heated body, and into the potential energy which those molecules acquire by reason of the expansion of the heated body. There are some operations, as, for example, the melting of a solid, which cannot be explained completely without additional assumptions. These assumptions, however, are not inconsistent with the kinetic theory, but extend it by considering the motion of the atoms as well as that of the molecules.

282. The Mechanical Measure of Temperature. — If we adopt the kinetic theory of heat, we are naturally led to consider the possibility of finding a measure of temperature in terms of the motions of the molecules of the hot body. In the case of a gas such a measure is easily found. By comparing the experimental formula for the ideal gas, $pv = RT$ (§ 261), with the formula deduced from the postulates of the kinetic theory, $pv = \frac{1}{3}\, mnu^2$ (§ 185), we find

$$RT = \tfrac{1}{3}\, mnu^2,$$

and conclude that the temperature on the scale of the ideal gas thermometer is proportional to the mean kinetic energy of the molecules of the gas. Since by the Maxwell-Boltzmann theorem (§ 188) the molecules of all gases, at the same temperature, have the same mean kinetic energy, this relation between energy and temperature is the same for all gases. We thus have a mechanical measure of temperature in the mean kinetic energy of the molecules of a gas.

Indeed the Maxwell-Boltzmann theorem goes further than this, and asserts that, when any assemblage of material particles is at a uniform temperature throughout, the mean kinetic energy which is associated with each degree of freedom of each particle is the same, and is proportional to the temperature measured on the scale of the ideal gas thermometer. If we use this theorem we can measure temperature in terms of a property which is common to all bodies at the same temperature.

The difficulties which arise when we apply this theorem to explain the peculiarities of the specific heats of bodies, and especially in connection with radiation, are so great, that it cannot be said

that the theorem is experimentally verified. It is nevertheless quite likely true that the mean kinetic energy associated with each degree of freedom of the particles — the atoms or molecules — of bodies at a common temperature is the same, and furnishes a measure of that property of those bodies which, considered in its relations to heat, is known as their temperature.

283. **The First Law of Thermodynamics.** — The subject of *thermodynamics* deals with the general laws governing the transformation of heat into work, and with the thermal properties of bodies which are the consequences of those laws. Its first postulate or law is that of the equivalence of heat and mechanical energy, and may be expressed, as before, by saying that heat is a form of energy, and that mechanical energy may be transformed into heat or heat into mechanical energy, without loss. If we represent by ΔQ a small amount of heat received by a body, by ΔE the increase occasioned thereby in the internal energy of the body, by ΔW the mechanical work done thereby by the body, in consequence of its expansion or otherwise, we may embody the first law of thermodynamics in the formula

$$\Delta Q = \Delta E + \Delta W.$$

Since this formula holds for all additions of heat to the body, it will hold also for the sum of all those additions, so that we may write

$$Q = E + W, \tag{102}$$

in which Q represents the algebraic sum of all the quantities of heat communicated to the body, and E and W represent the total gain in internal energy and the positive work done by the body.

284. **Cyclic Operations.** — In our study of the transformation of heat into work we are not in general able to make immediate use of the relation embodied in the last formula, because of our ignorance of what constitutes or measures the internal energy of a body. We avoid the complications arising from this circumstance by effecting the transformation by such a series of operations that the body employed, and all other parts of the system, return at the end of the series of operations to the same condition as that in which they were at the beginning. In this case we know that the internal energy of the body is the same at the end of the process as it was at the beginning, and hence that E in Equation 102 equals 0, and so $Q = W$. The excess of heat received by the body over that emitted by it is equal to the work done by it in the process.

The body used to receive and emit heat and to do work is called

the *working body*. A series of operations by which the working body is brought back to its original state is called a *cyclic operation*, or a *cycle*. By using only cyclic operations we can study the transformation of heat into mechanical energy without knowing anything very definite about the properties of the working body.

The use of cyclic operations for this purpose was introduced in 1824 by Sadi Carnot.

285. Isothermals. — The working body which we shall employ is any body which expands as its temperature rises, and does work by expanding against uniform hydrostatic pressure. The temperature of such a body falls when the body expands without receiving heat, and rises when the body is compressed without being allowed to emit heat. To keep the body at a constant temperature while it is expanded or compressed, heat must be allowed to enter it or leave it.

A body of which the volume changes while its temperature is kept constant is said to experience an isothermal expansion or contraction. It is convenient to represent such an isothermal change by a line in a diagram, constructed by using the corresponding values of pressure and volume, which hold for the working body while its temperature is kept constant, as ordinates and abscissas with reference to rectangular axes of pressure and volume (Fig. 145).

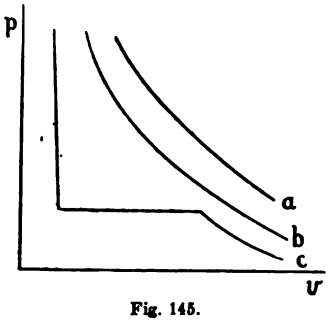


Fig. 145.

Such a curve is called an isothermal curve, or line, or simply an *isothermal*. Without particularizing more than we have done about the properties of the working body, we are able to determine certain general features of all isothermals.

No two isothermals can intersect, for their intersection would imply that the working body could assume the same volume under the same pressure while in two different conditions as regards temperature; or that, otherwise considered, the temperature of the working body might be changed without changing either its pressure or volume.

Isothermals are generally so shaped that smaller pressures correspond to larger volumes. They generally slope downward from left to right, as represented in lines *a* and *b* of the diagram. This condition represents the general condition necessary for the maintenance in equilibrium of a body under pressure. In the special

case of a body existing in two states at once the isothermal has a horizontal portion, as represented in line *c* of the diagram.

The isothermals drawn in Fig. 145 may not represent the true isothermal relations of pressure and volume for any real body, but they do correctly represent the isothermal relations of all bodies of the sort we have chosen to serve as working bodies.

286. Adiabatics. — A body of which the volume changes while it is so situated that it can neither receive nor emit heat is said to experience an adiabatic expansion or contraction. Adiabatic changes can be represented in a diagram by curves drawn to represent the corresponding values of pressure and volume when the volume of the body is changed without allowing it to receive or emit heat. Such a curve is called an adiabatic curve or line, or simply an *adiabatic*.

Just as with the isothermals, we are able to determine certain general properties of the adiabatics of the working bodies which we have chosen to employ.

No two adiabatics can intersect, for their intersection would imply that a body might receive heat without experiencing any change in its volume, pressure, or temperature.

Adiabatics always are so shaped that smaller pressures correspond to larger volumes. They always slope downward from left to right. This is necessary in order that the body under pressure shall be in equilibrium. Even in the case of a body existing in two states at once, an adiabatic change in volume will occasion a change in temperature and a consequent change in pressure. The adiabatics of such a body have no such horizontal portions as its isothermals have.

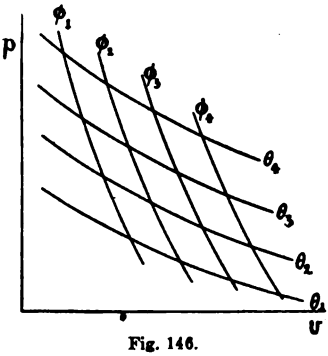


Fig. 146.

The slope of the adiabatic which is drawn through any point on the diagram is always steeper than that of the isothermal drawn through the same point. To perceive this, consider any small change of volume, say a contraction. If this change occurs isothermally, heat is removed from the body as its pressure rises. If an equal change in volume occurs adiabatically, the heat produced by compression is not removed, the temperature rises, and the pressure attained is consequently higher than that attained by an equal isothermal contraction.

In Fig. 146 the curves marked θ_1, θ_2, etc., are typical isothermals; the curves marked ϕ_1, ϕ_2, etc., are typical adiabatics.

287. Conditions for Reversibility. — If a body is placed in an envelope impermeable to heat, and with pressure and volume defined, so that it is in equilibrium, an infinitesimal change in the pressure will occasion an adiabatic change of volume, but will not cause the body to depart finitely from its condition of equilibrium; so that a body may be made to describe, as is said for brevity, an adiabatic line without ever deviating finitely from a condition of equilibrium. It may be made to retrace that line by a series of infinitesimal changes in which the pressures will not differ finitely from those to which the body was subjected at the same volumes during the direct process. The process may be stopped at any point and leave the body in mechanical and thermal equilibrium.

When the body is required to describe an isothermal, it is necessary, in order that it may always remain in thermal equilibrium, to provide a reservoir of heat with which it may exchange heat. The temperature of this reservoir must be that corresponding to the isothermal to be described. As the pressure is changed the temperature of the body changes, but its change never becomes finite so long as the body receives heat from the reservoir or emits heat to it. The body may therefore be made to describe an isothermal without deviating finitely from the conditions of mechanical and thermal equilibrium. It may also retrace the isothermal by a series of changes in which its pressures and its temperatures will not differ finitely from those which were experienced by the body at the same volumes during the direct process. The process may be stopped at any point and leave the body in mechanical and thermal equilibrium.

We may sum up by saying that the condition that any change of the body shall be reversible is that the pressure upon it shall never be discontinuously changed; and further in the case of the isothermal change, that the temperature of the reservoir shall never be finitely different from that of the body.

288. Work Done by Expanding Body. — When a body expands under the conditions of equilibrium, it does work. In the simplest case, in which the expansion takes place under constant pressure, the work done is equal to the product of the pressure and the increase of volume. To show this we consider the body contained in a cylinder (Fig. 147) and confined by a piston placed at C and maintained in position by a force F. If we represent the pressure

of the body by p and the area of the face of the piston by a, we may write (§ 119) $F = pa$. Now suppose the piston to move out through the distance s, to C'; then the work done against the force F by the expanding body is $Fs = pas$. But as measures the increase of the volume of the body by the expansion, so that the work done is equal to the product of the pressure and the increase of volume.

In case the increase in volume is accompanied by a change in pressure we may represent the work done by the aid of a diagram without attempting to calculate its actual amount. We represent the successive pressures and volumes by a line referred to axes of pressure and volume (Fig. 148).

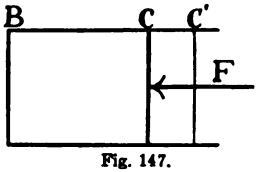


Fig. 147.

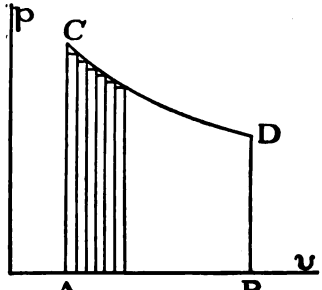


Fig. 148.

The points C and D at the ends of this line represent the initial and terminal conditions of the body. The distance from A to B represents the whole increase of volume. Divide this distance up into a great number of small portions. Each one of these represents an increase in volume within which the pressure, represented by the ordinates of the curve erected on the points of this portion, is approximately constant. The product of this pressure by this small increase in volume is approximately equal to the work done during the small expansion, and is represented approximately on the diagram by the area of a rectangle whose base is the small distance representing the small expansion and whose height is the pressure exerted during that small expansion. The sum of all such rectangles will represent approximately the work done by the whole expansion, and as the subdivision of the line AB is made more minute, this approximation becomes closer. In the limit the sum of the rectangles is equal to the area included under the curve CD between the perpendiculars representing the initial and final pressures. This area, therefore, represents the work done by the expansion.

In case the line representing the expansion is straight or is so small a portion of a curved line that it may be taken as straight to a first approximation (Fig. 149), the work done by such an expansion is given by the formula

$$W = \frac{p_1 + p_2}{2}(v_2 - v_1).$$

When the change of volume is an expansion, the work done by the body is positive; when it is a contraction, the work done by the body is negative, or positive work is done on the body.

When the body performs a cycle and returns to its initial condition, the work it has done is represented by the difference between the areas representing the work done by it and the work done on it. This difference (Fig. 150) is simply the area included in the

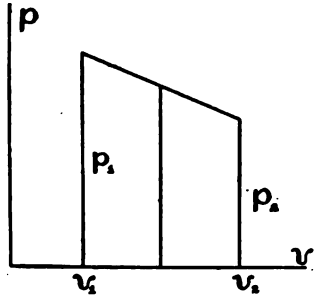


Fig. 149.

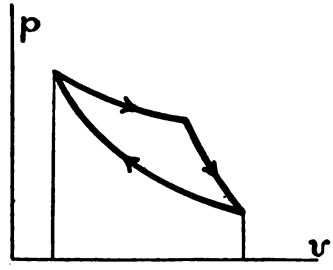


Fig. 150.

lines representing the cycle, and is positive if the work done by the body is represented by the larger area. As the diagram shows, it is positive if a person, passing around the cycle in the sense in which it was described, has the area included by it on his right hand.

289. Carnot's Cycle. — We are now ready for the consideration of a peculiarly simple cycle, introduced by Carnot, and known as *Carnot's cycle*, by the aid of which the fundamental principles of thermodynamics may be discussed. It is a cycle consisting of two isothermal portions, and two adiabatic portions (Fig. 151), as the diagram shows. It is commonly considered as the cycle performed by an ideal engine, called a *Carnot's engine*, doing work by the aid of heat drawn from a *source* at the temperature θ_2 of the higher isothermal, and delivering heat to a *refrigerator* at the temperature θ_1 of the lower isothermal.

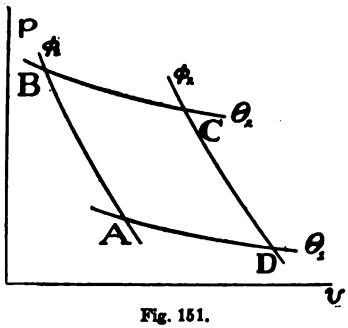


Fig. 151.

Starting with the working body in the condition represented by the point A (Fig. 151), we may describe the four steps of the Carnot's cycle as follows:

1. An adiabatic compression from A to B, by which the temperature of the working body is raised from θ_1 to θ_2.

2. An isothermal expansion from B to any point C, during which the working body receives the quantity of heat Q_2 from the source.

3. An adiabatic expansion from C to D, by which the temperature of the working body is lowered from θ_2 to θ_1.

4. An isothermal compression from D to A, during which the working body gives up the quantity of heat Q_1 to the refrigerator.

By the performance of this cycle, work has been done by the body, represented by W and measured by the area $ABCDA$. Heat has been received from the source, and given out to the refrigerator; and since the condition of the working body is the same at the end of the operation as at the beginning, its internal energy is also the same, and we may write (§ 284)

$$Q_2 - Q_1 = W.$$

Since each step in this cycle is reversible, the cycle itself is reversible. When it is performed in the reverse way, the work W is done on the body, heat equal to Q_1 is taken from the refrigerator, and heat equal to Q_2 is delivered to the source. The equation $Q_2 - Q_1 = W$ still holds.

290. Carnot's Theorem. — The combination of a working body with suitable sources and refrigerators is called a *thermodynamic engine*. The simplest combination of this sort performs Carnot's cycle and is called Carnot's engine.

The *efficiency* of a thermodynamic engine is defined as the ratio between the work done by it and the heat taken from the source. By considering a cycle similar to that which we have described, Carnot convinced himself that the efficiency of an engine performing a reversible cycle is at least as great as that of an engine performing any cycle in which it exchanges heat with the same source and refrigerator. This statement is known as *Carnot's theorem*, although Carnot's proof of it, which involved the older notion of the materiality and indestructibility of heat, is not tenable in the light of our present knowledge. The theorem seemed to Clausius (1850) and to Lord Kelvin (1851) so likely to be true that they independently undertook to investigate the conditions which would have to be assumed in order that it might be proved true. Their investigations led them to the discovery of a principle which is known as the second law of thermodynamics.

To lead up to this principle let us consider a cyclic operation

which is performed by two engines, one of which is reversible, while the other may or may not be. For convenience in designating the latter, we shall call it the irreversible engine. We assume, for the sake of argument, that the irreversible engine is more efficient than the reversible one, and we gear the two engines together, so that the irreversible engine, working forward, spends all its work in driving the reversible engine backward. The irreversible engine takes a quantity of heat Q_2' from the source and delivers a quantity of heat Q_1' to the refrigerator. The work W which it does is equal to the heat transformed by it, so that we have

$$W = Q_2' - Q_1'.$$

This work is used in driving the reversible engine backward, taking from the refrigerator the quantity of heat Q_1 and delivering to the source the quantity of heat Q_2. The work done is again equal to the difference between these quantities of heat, so that we have

$$W = Q_2 - Q_1.$$

Now if the reversible engine were to run forward, and to do the work W, it would take from the source the quantity of heat Q_2. Its efficiency is therefore given by W/Q_2, and this by hypothesis is less than the efficiency W/Q_2' of the irreversible engine. On this hypothesis, therefore, we have $Q_2' < Q_2$, and since

$$W = Q_2' - Q_1' = Q_2 - Q_1,$$

we have also $Q_1' < Q_1$.

Writing the equation above in the form $Q_2 - Q_2' = Q_1 - Q_1'$, we obtain two equal terms, one of which represents the gain in heat of the source, which is positive, since $Q_2 > Q_2'$, the other the loss of heat of the refrigerator, which is necessarily also positive. On the whole, no external work has been done by the system of engines, and the final result of the operation, which can be repeated as many times as we please, is that no work is done and that heat is transferred from the refrigerator to the source. In other words, if an irreversible engine, or indeed any engine, were more efficient than a reversible engine, it would be possible, without expending any energy, and without any essential change occurring in the system, to transfer heat from the colder of two bodies to the hotter one. This conclusion is contrary to all our experience. By denying its correctness in fact, we are thus led to deny the validity of the hypothesis from which it is derived, and so to assert that the efficiency of a reversible engine is at least as great as that of any engine using the same source and refrigerator.

291. The Second Law of Thermodynamics. — The principle which we have been led to adopt, in order to demonstrate Carnot's theorem, was stated by Clausius in a more formal way in the words: Heat cannot pass of itself (or without compensation) from a colder to a hotter body. This statement is one of the forms assumed by the second law of thermodynamics.

292. Absolute Scale of Temperature. — It is easy to extend the demonstration just given so as to show that the efficiency of all reversible engines, which work between the same source and refrigerator, is the same. The efficiency of the reversible engine is therefore independent of the particular material used in the working body. It depends only on the temperatures of the source and the refrigerator. In other words, the efficiency of the Carnot's engine is a function of the temperatures of the source and the refrigerator.

In the definition of the scales of temperature which we have hitherto used, we avail ourselves of some property of a body which is a function of its temperature. For example, in our definition of the Centigrade scale we make use of the volume of mercury contained in glass. In this instance, as in all the others, the property employed is that of a particular body, or at best, of a particular substance.

We have now discovered, by the study of Carnot's cycle, a property, depending only on temperature, which is not peculiar to any one body or to any one substance, but is common to all. We may therefore use this property as a means of defining a scale of temperature in such a way as to make our definition independent of any particular substance. The scale which we shall thus define is called *the absolute scale of temperature.*

A complete definition of the absolute scale is obtained by the choice, first, of the form of the temperature function which measures the efficiency, and secondly, by a specification of the length of the degree, or of the numerical value of some standard temperature, on that scale.

One general remark may here be made, which holds true whatever be the form of the temperature function chosen. The efficiency of an engine working from any given source will increase as the temperature of the refrigerator falls; but it can never exceed unity, for it is manifestly impossible for more work to be done by the engine than is equivalent to the heat received by it from the source. The temperature of the refrigerator for which the efficiency becomes unity is therefore the lowest conceivable temperature. There can

be no lower temperature than that, and it must be the same in all bodies which will serve, when used as refrigerators, to make the efficiency equal to unity. This temperature is therefore a natural zero, and it is advisable, in constructing the absolute scale, to mark it as the zero point of that scale. All actual temperatures will then be positive.

Turning now to the choice of the particular temperature function by which the efficiency shall be expressed, we are guided in making it by the study of the efficiency of an engine in which the ideal gas is used as the working body. If we express the temperatures of the source and the refrigerator on the scale of the ideal gas thermometer, we find that the efficiency of such an engine is equal to the difference between the temperatures of the source and of the refrigerator, divided by the temperature of the source. That is, if we represent these temperatures by S and R respectively, the efficiency is expressed in terms of them by the formula

$$\frac{Q_2 - Q_1}{Q_2} = \frac{S - R}{S}. \tag{103}$$

Since the efficiency does not depend on the nature of the working body, this expression is also the measure of the efficiency when any other substance than the ideal gas is used as the working body. This relation so commends itself by its simplicity that we adopt it as the general relation, or form of the temperature function, by which the efficiency is expressed. That is, we suppose the efficiency to be known, and define the temperatures of the source and the refrigerator by a formula similar in form to that just given.

From this formula it appears that the efficiency of the engine becomes unity when the temperature of the refrigerator becomes the zero of the ideal gas thermometer. This zero is therefore the lowest conceivable temperature, and we may adopt it as *the absolute zero*, or zero of the absolute scale.

It remains to adopt the length of a degree on the absolute scale. It is convenient to have this length as nearly as possible the same as the length of the Centigrade degree. We accordingly adopt 100 degrees as the difference of absolute temperature between the standard points of the Centigrade scale. By calculation from the known properties of gases we find that the efficiency of an engine working between these temperatures is 100/373, and hence, putting $\frac{S - R}{S} = \frac{100}{373}$, and $S - R = 100$, we obtain 373 for the absolute tem-

perature of boiling water and 273 for the absolute temperature of melting ice. After the numerical value of one temperature is fixed, the numerical value of any other temperature is found by determining the efficiency of an engine working between the two temperatures, and using the general formula.

We have thus obtained a complete scale of temperature, starting from the absolute zero, and determined in a way which is independent of the characteristics of any particular substance. It is the same as the scale of the ideal gas thermometer, but has not the hypothetical character of that scale. For all ordinary purposes a close approximation to the absolute temperature of a body may be obtained by adding 273 to its temperature on the Centigrade scale.

293. Entropy.—From the equation of § 292 defining the absolute scale of temperature, $\frac{Q_2 - Q_1}{Q_2} = \frac{S - R}{S}$, we obtain $\frac{Q_2}{Q_1} = \frac{S}{R}$, and may define the absolute temperature in another, not essentially different, way, by saying that in the description of Carnot's cycle the heat taken from the source is to the heat given out to the refrigerator as the absolute temperature of the source is to that of the refrigerator.

From this relation we have further, $\frac{Q_2}{S} = \frac{Q_1}{R}$, in which the quantities relating to source and refrigerator are separated in two similar terms. Either of these quantities defines a change in a certain quantity, called by Clausius the *entropy*, of which all that we can say in this place is that it is a function of the condition of the body under consideration, which remains constant when the body undergoes an adiabatic change and increases as the body undergoes an isothermal change by an amount expressed by the ratio of the heat which enters the body to the absolute temperature of the body. The various adiabatics will be characterized by different values of this function. They are lines of equal entropy, and are therefore often called *isentropics*.

The entropy of a system of bodies is the sum of the entropies of its parts.

As may be seen at once from the equality of the two ratios which appear in the description of Carnot's cycle, the entropy of the system is not changed by the description of the cycle, for the working body returns to its original state, and the entropy lost by the source is equal to that gained by the refrigerator. This unchange-

ableness of the entropy in the case of the simple cycle will hold also for all more complicated reversible cycles, provided that they are capable of analysis into simple cycles. This amounts in effect to saying that the entropy remains unchanged by the performance of reversible cyclic operations.

In many cases in which the cyclic operations performed are in some part irreversible, it may be shown that the result is to increase the entropy of the system. While it is perhaps not possible to give a general and convincing proof that the entropy must always be increased by irreversible operations, yet that this rule or law is true is generally accepted as proved by the complete success with which it has been applied to the solution of various thermodynamic problems. Used in this way, it is called the *principle of the growth of entropy*. It may be considered as a generalization from our experience of the same sort as the second law of thermodynamics, and ranking as another and more general statement of it.

The entropy is a quantity which is of equal importance with the energy in characterizing the condition of a system of bodies. The two principles of the conservation of energy and of the growth of entropy were generalized by Clausius in the statements:

The energy of the universe is constant.

The entropy of the universe tends toward a maximum.

294. Thermodynamic Properties of Bodies. — When the properties of bodies are examined by the methods of thermodynamics, many important relations are found among them, which are confirmed by experiment. For example, it may be proved, as a general principle, that a substance which expands under constant pressure when its temperature rises has its temperature raised by an adiabatic compression; and that a substance which contracts when its temperature rises has its temperature lowered by adiabatic compression. This general conclusion was confirmed by Joule for the case of water. In the interval between 0° and 4° C., within which water contracts as its temperature rises, a sudden compression will lower its temperature. At any temperature higher than 4°, a sudden compression will raise its temperature.

Again, it may be proved, as a general principle, that the melting temperature of a substance, that is, the temperature at which the solid and liquid states of the substance are in equilibrium, will rise if the pressure is increased, provided the relative density of the solid is greater than that of the liquid; but it will fall if the pressure is increased, provided the relative density of the solid is less than

that of the liquid. This principle was confirmed for the case of water and ice by the experiments of James Thomson and his brother, Lord Kelvin. Water is a substance belonging to the second class, in which the density of the solid is less than that of the liquid. When a mixture of water and ice, whose temperature under atmospheric pressure is 0° C., was subjected to additional pressure, its temperature fell. The temperature rose again when the pressure was removed. By using this principle we explain the melting of ice under pressure, and its return to the solid state, or its regelation, when the pressure is removed. Other substances, for example, paraffine, which belong to the first class, were shown by Bunsen and by other observers to exhibit a rise of temperature when pressure was applied to them.

The methods of thermodynamics have been applied, not only to the study of the general properties of bodies, but more particularly to the study of solutions, and of mixtures of different substances which act on each other chemically. In this way a foundation has been laid for the physical study of chemical action.

295. Dissipation of Energy. — According to the general principle of conservation, the energy of the universe is constant. Not all of it, however, is available for doing work. In all cases of the transformation of energy in which nonconservative forces act, some at least of the energy appears in the form of heat. If such transformations were continued long enough, all other energy in the universe would be transformed into heat, and if this heat could be transformed into mechanical energy, all the energy of the universe would be available for use as mechanical energy. But this is not the case. We know, from our study of Carnot's engine, that even when heat is transformed into work in the most efficient way possible, some of the heat is not transformed, but is transferred from a higher to a lower temperature. None of the heat in the body at lower temperature can be utilized, unless a still colder body can be found to serve as refrigerator. It is plain, therefore, that the final effect of all transformations of energy will be to bring about a common temperature of all bodies. When this common temperature has been attained, no further use can be made of the heat in those bodies, and since, by hypothesis, all other available energy has been turned into heat, no available energy will be left. The amount of energy in the universe will be still the same, but none of it can be used.

The loss of available energy goes on more rapidly than it would

if it were possible to utilize only reversible processes for the transformation of heat into work. Whenever an irreversible process takes place, — and such processes are continually taking place in nature, — some of the energy which might have been utilized to do work is rendered unavailable. It is said to be dissipated, and the process by which it is rendered unavailable is called the *dissipation of energy*.

Boltzmann has explained the second law of thermodynamics by the aid of the kinetic theory of heat. On the assumption that the atoms of all bodies are in unordered motion, he has shown that the most probable condition of any assemblage of atoms is one in which the mean kinetic energy of each atom is the same, and that, when this condition does not exist, the probable change in the assemblage is toward this condition, and not away from it. He considers that the second law of thermodynamics is the experimental equivalent of this theoretical conclusion, and expresses the general tendency of bodies toward this most probable condition. Boltzmann has suggested a possible escape from the conclusion that the final condition of the universe is that in which there is no available energy, by calling attention to the fact that, although the general tendency of any system is toward its most probable condition, yet there may occur in the system such a combination of circumstances that for a while the tendency will be toward an improbable condition. During this period the second law will not hold true, and at the end of it, the system will have acquired a store of available energy.

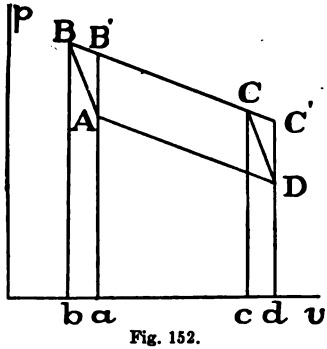


Fig. 152.

EXAMPLES, XV

1. *To calculate the efficiency of a Carnot's engine, in which an ideal gas is the working body.*

We may calculate the efficiency in this case by considering a cycle so small that its component elements may be taken as straight, and the cycle itself a parallelogram (Fig. 152). The performance of this cycle involves doing positive work by expansion along one adiabatic, and negative work by contraction along another adiabatic. The temperature changes in the first case from θ_2 to θ_1, and in the second case from θ_1 to θ_2. It is easy to show that the two amounts of work done by these changes are equal and cancel each other.

To do this let us consider the equivalent statement, that the work done by any adiabatic expansion of a gas, by which its temperature falls from the temperature θ_2 to the temperature θ_1, is the same. When a gas expands adiabatically, since no heat is received from without, the work which it does is done at the

expense of its internal energy. But by Joule's law (§ 265) the internal energy of a gas depends only on its temperature, and hence the change of its internal energy will be the same whenever the change of temperature is the same. When the temperature of the gas falls by adiabatic expansion from θ_2 to θ_1 it loses the same amount of internal energy, on whatever adiabatic the expansion takes place, and so does the same amount of external work by any such expansion. For this reason, in studying the cycle of the ideal gas engine we do not need to calculate the work done by the adiabatic operations.

Again, by Joule's law (§ 265) the internal energy of the gas does not change if its temperature does not change, so that the heat received or emitted during an isothermal operation is equal to the work done during that operation.

After these preliminary remarks we may turn to the consideration of our problem. As was shown in the last paragraph, we can set the heat Q_2 taken from the source by the isothermal expansion from B to C (Fig. 152) equal to the work done by that expansion, represented by the area under BC. Similarly the heat Q_1, emitted during the isothermal compression from D to A, is equal to the work done on the gas during that compression, represented by the area under AD. We know that the work W done in the cycle is represented by the area $ABCD$.

The efficiency

$$\frac{W}{Q_2} = \frac{Q_2 - Q_1}{Q_2} = 1 - \frac{Q_1}{Q_2}$$

can be now calculated by obtaining the ratio of the areas which are equal to Q_1 and Q_2.

Since $Q_2 = W + Q_1$, and since W = area $ABCD$ = area $AB'C'D$, we have Q_2 = area under $B'C'$.

Hence

$$\text{efficiency} = 1 - \frac{\text{area under } AD}{\text{area under } B'C'}.$$

These areas are to each other as the mean values of their ordinates, which represent the pressures of the gas at the two temperatures θ_1 and θ_2; and from the formula expressing the Boyle-Gay-Lussac law (§ 261) these pressures are proportional to the temperatures θ_1 and θ_2 measured on the scale of the ideal gas thermometer. Therefore we have

$$\frac{\text{area under } AD}{\text{area under } B'C'} = \frac{\theta_1}{\theta_2},$$

and

$$\text{efficiency} = 1 - \frac{\theta_1}{\theta_2} = \frac{\theta_2 - \theta_1}{\theta_2}.$$

2. *To discuss the efficiency of the reversible engine on the assumption of the materiality of heat.*

As in § 290 we assume a reversible engine working backward, driven by another engine, called irreversible, working forward. Assume the irreversible engine to be more efficient than the reversible one. Then if the irreversible engine takes heat Q from the source and transfers it to the refrigerator, it will do work W', which is greater than the work W which must be done on the reversible engine to transfer the heat Q back from the refrigerator to the source. By driving the irreversible engine forward, and the reversible engine backward in this way, the source and the refrigerator neither gain nor lose heat, and with every cycle described surplus work equal to $W' - W$ will be done.

Carnot reasoned in this way, and concluded, on the basis of the maxim that "the perpetual motion is impossible," that the necessary result was inconsistent with natural law. He therefore concluded that the efficiency of a reversible engine could not be less than that of any engine.

3. *To determine the necessary consequences if the two oppositely working engines of* § 290 *are so controlled that the irreversible engine takes from the source the same amount of heat as the reversible engine returns to it.*

Suppose that the irreversible engine takes heat Q_2 from the source, does work W', and renders heat Q_1' to the refrigerator. The reversible engine, worked backwards, takes heat Q_1 from the refrigerator, has work W done on it, and renders heat Q_2 to the source. Since by hypothesis the irreversible engine is the more efficient one, we have $\frac{W'}{Q_2} > \frac{W}{Q_2}$, or $W' > W$; and since $Q_2 - Q_1' = W'$, $Q_2 - Q_1 = W$, we have $Q_1 > Q_1'$. The source neither gains nor loses heat, surplus work equal to $W' - W$ can be obtained in each cycle, and the refrigerator loses heat equal to $Q_1 - Q_1'$; so that work can be got by cooling off the colder of the two bodies which serve as sources of heat. This conclusion seems so improbable that we refuse to accept it, and so deny the hypothesis from which it is derived.

This is Lord Kelvin's form of the proof of Carnot's theorem. Lord Kelvin stated the second law of thermodynamics, in conformity with this form of proof, by saying: It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects.

4. *To show that the entropy of a system of two bodies, at the absolute temperatures T_1 and T_2, of which T_1 is higher, is increased by the passage of heat by radiation or conduction from one to the other.*

The entropy lost by the hotter body when the heat Q leaves it is Q/T_1; that gained by the colder body when it receives this heat is Q/T_2. The total gain in entropy is $\frac{Q}{T_2} - \frac{Q}{T_1} = Q\frac{T_1 - T_2}{T_1 T_2}$, and this is positive, since $T_1 > T_2$.

The heat in this case is transferred by an irreversible operation. The result is typical of all irreversible operations in a closed system, in which the entropy always increases.

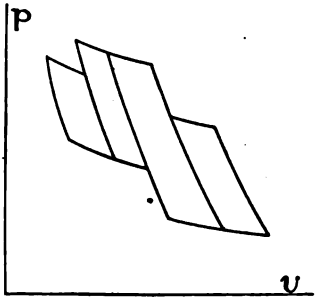


Fig. 153.

5. *To examine cycles which are reversible but not simple.*

Let there be a cycle consisting of any number of alternate isothermal and adiabatic portions (Fig. 153). By prolonging the adiabatics as indicated, this may be cut up into a number of Carnot's cycles, to each one of which a formula like $\frac{Q_2}{S} - \frac{Q_1}{R} = 0$ will apply.

To generalize this result it is convenient to make the convention that all signs shall be written positive, and that in effecting any summation indicated, heat taken in by the working body shall be considered positive, heat given out, negative. Represent by Q_1 the heat taken in, positively or negatively, by one of the isothermal parts of one of these cycles, and by T_1 the absolute temperature of the isothermal. Then for all the cycles we shall have

$$\frac{Q_1}{T_1}+\frac{Q_2}{T_2}+\frac{Q_3}{T_3}+\cdots\frac{Q_n}{T_n}=0, \text{ or } \sum\frac{Q}{T}=0.$$

6. *To find the peculiarity of the sum* $\frac{Q_1}{T_1}+\frac{Q_2}{T_2}$ *when the cycle to which it refers is irreversible.*

The efficiency of the irreversible cycle is less than that of a reversible one, so that $\frac{Q_2-Q_1}{Q_2}<\frac{T_2-T_1}{T_2}$. This may be represented in an equation by writing $\frac{Q_2-Q_1}{Q_2}=\frac{T_2-nT_1}{T_2}$, using n to represent some factor larger than 1. From this we get $\frac{Q_2}{T_2}-\frac{Q_1}{nT_1}=0$. In this equation, for the sake of greater clearness, the signs referring to heat taken in and given out by the working body are explicitly retained. Since n is greater than 1, the negative term Q_1/nT_1 is numerically less than Q_1/T_1, so that the sum $\frac{Q_2}{T_2}-\frac{Q_1}{T_1}$ will be negative. Using now the convention about the signs explained in Example 5 we may write for the case of the irreversible cycle,

$$\frac{Q_1}{T_1}+\frac{Q_2}{T_2}<0.$$

Since a similar relation will hold for any part of a more complicated cycle, like that of Example 5, which is irreversible, we may set for any cycle that is in whole or in part irreversible $\Sigma Q/T<0$.

7. *To show that the entropy of a closed system increases when an irreversible operation takes place in it.*

Suppose that the system is in a state represented by A and changes irreversibly to a state represented by B. It can be brought back to the state A by a suitable series of reversible operations, and to the cycle thus completed the inequality of Example 6 will apply. Now if we represent by ϕ_A and ϕ_B the entropies of the system in the states A and B, we have from § 293, $\phi_A-\phi_B=\sum_B^A\frac{Q}{T}$, in which the summation is taken over the reversible part of the cycle. For the whole cycle $\Sigma Q/T<0$, and if we use $\sum_A^B Q/T$ to represent the summation taken over the irreversible part of the cycle, we have

$$\sum\frac{Q}{T}=\sum_A^B\frac{Q}{T}+\sum_B^A\frac{Q}{T}=\sum_A^B\frac{Q}{T}+\phi_A-\phi_B<0.$$

Now the system is supposed to be closed, that is, it is supposed neither to receive nor to emit heat during the irreversible operation. On this supposition $\sum_A^B\frac{Q}{T}=0$, and hence $\phi_A-\phi_B<0$, or $\phi_B>\phi_A$; the entropy in the state reached by the irreversible operation is greater than that in the initial state.

This is a sketch of the argument by which Clausius was led to the conclusion that, since most natural operations are irreversible, the entropy of the universe tends to a maximum.

8. *To construct a diagram to represent a Carnot's cycle with temperatures as ordinates and entropies as abscissas.*

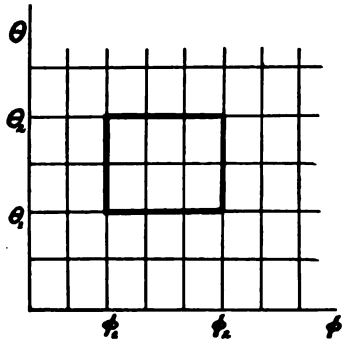


Fig. 154.

The isothermal and adiabatic (or isentropic) steps of the cycle are represented by lines of equal length parallel to the axes, so that they enclose a rectangle (Fig. 154). The heat Q_2 is taken in at the temperature θ_2 and we have $Q_2 = (\phi_2 - \phi_1)\theta_2$. Similarly $Q_1 = (\phi_2 - \phi_1)\theta_1$, and hence

$$Q_2 - Q_1 = (\phi_2 - \phi_1)(\theta_2 - \theta_1).$$

But $Q_2 - Q_1 = W$, the work done during the cycle, so that the area enclosed by the isothermal and adiabatic lines measures the work done, as in the ordinary representation of Carnot's cycle (§ 289).

BOOK V

LIGHT

CHAPTER XVI

GEOMETRICAL OPTICS

296. General Considerations Respecting Light. — The sense of sight presents to us a large number of phenomena which we associate with the general idea of *light*. Our object is to study and classify these phenomena, and so far as possible to explain them. We shall find in this case, for the first time, that a sufficient explanation cannot be given in terms of the motion of matter, and that it will be necessary to extend our conception of the physical universe by assuming the existence of an additional constituent of it as the medium in which the actions take place which constitute light. We shall find further that this medium has properties unlike those of any known body, and that the actions in it do not conform to the laws of action of any known body. Our explanation therefore consists finally in the reduction of all the phenomena of light to actions which take place in a hypothetical medium according to certain formal laws.

297. Origin of Light. — Opinions were divided, in antiquity, between the view that light was an affection of external bodies, passing from them to the eye, and the view that the eye itself was the source of light, which passed out from it to any object at which the eye was looking. The latter view was unable satisfactorily to meet the criticism of Aristotle, to the effect that, if it were true, we should be able to see in the dark. It nevertheless held a prominent place in thinking for many years after that criticism was made on it. Gradually, however, the other view displaced it.

298. Rectilinear Transmission of Light. — The earliest observations must have shown that light travels in straight lines. This fact is so well known that it does not need illustration by special experiments. In our subsequent discussions we shall use the word *ray* to denote any straight line in which light travels.

299. Reflection of Light. — Another observation, which was made so early that no record of it has come down to us, is that of the *reflection* of light. When light, in passing from an object, encounters a smooth surface, it is more or less completely turned back or reflected from that surface. If we select a point on the reflecting surface (Fig. 155) and draw from it a normal *ON* to that surface, and if we then draw from the source of light the ray *SO* to that point, this ray will make an angle α with the normal, which we call the *angle of incidence*. The angle α' made by the reflected ray *OP* with the normal is called the *angle of reflection*. The light from the source is reflected at the point of incidence according to the following law: The incident and reflected rays lie in a plane which contains the normal to the reflecting surface, and the angles of incidence and reflection are equal.

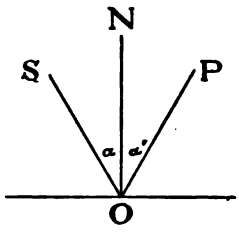


Fig. 155.

When light falls upon a rough or irregular surface, the various small surfaces of the body reflect the light in different directions, so that it reaches an observer wherever he may be. Such reflection is called irregular or *diffuse reflection*.

300. Refraction of Light. — A third observation, which for the present we shall describe only in general terms, is that of the *refraction* of light. This refraction is the sudden bending or change of direction which a ray of light exhibits as it passes from one medium into another. It is refraction which makes a straight stick look bent when one end of it is thrust into water.

301. Sources of Light. — Many bodies give forth light of themselves, so that if one of them is brought into a perfectly dark room, it will be visible. Such bodies are called *self-luminous* bodies. Other bodies, which are not self-luminous, are seen by the light which falls upon them from self-luminous bodies, and is irregularly reflected by them. For use in the study of geometrical optics, any body, from which light comes in sufficient intensity to render it visible, may be taken as a *source* of light, though, strictly speaking, the only true sources of light are self-luminous bodies. A source which is so small that for the purposes of the discussion it may be considered a point we shall call a *point source*. An extended source is called an *object*.

The light from a point source proceeds outwards in diverging rays. If a small area is taken in space, it will generally limit a cone

of rays, whose apex is at the source, and of small angular aperture. Such a cone of rays is called a *pencil* of light. A similar limited assemblage of rays, considered simply as rays and without reference to their origin, is called a *beam* of light.

The eye perceives a source of light by receiving from it a pencil of light, which is limited by the pupil of the eye.

An observer judges the source to lie on the axis of the pencil, and even with one eye is able to form some judgment as to how far away it lies, by an unconscious estimate of the divergence of the rays which form the pencil. If the pencil enters the eye after reflection from a mirror or refraction through a lens, the immediate judgment of the observer places the source on the axis of the pencil, just as if it had entered the eye directly from the source.

When the same source is observed by both eyes, the axes of the pencils received intersect at the source, and by an unconscious estimate of the divergence of these axes, an estimate of the distance of the source can be made, which is far more precise than any that can be made with one eye.

302. Intensity of Light. — By the *intensity* of light is meant its illuminating power, as judged by the eye on observation of the illumination of some standard surface. It is measured, or rather two intensities are compared, by an instrument called the *photometer*.

One of the earliest forms of photometer, invented by Count Rumford, is made by setting up a vertical rod at a little distance from a white screen. The two sources of light to be compared are set so as to cast shadows of the rod near each other on the screen. The space covered by the shadow from one source is illuminated by the light from the other source. The sources are then moved about until the two shadows appear equally illuminated, and it is then said that the intensity of the light from the two sources is the same. By comparing the effects of the sources when they are set at different distances from the screen, it is found that the intensity of the light from a source varies inversely with the square of the distance from the source.

303. Geometrical Optics. — If we use the observed facts which have been described, and represent rays of light by lines, it is evident that we may apply the methods of geometry to the study of the reflection and refraction of rays of light by variously shaped bodies, and that for this study it is not necessary for us to know anything about the nature of light. The limited subject which is treated in this way is called *geometrical optics*. In general, geometrical optics is concerned with the study of mirrors, lenses, and prisms, and of the optical instruments constructed by their combinations.

Reflection of Light. Mirrors.

304. The Plane Mirror. — A mirror is, in general, a polished surface used for the reflection of light. The surface usually conforms very nearly to some geometrical surface. The most common mirrors have either plane or spherical surfaces.

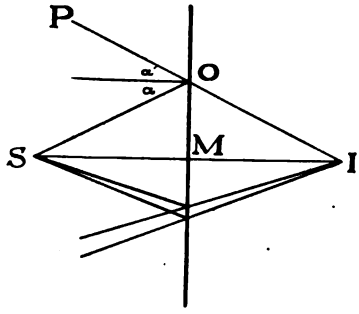


Fig. 156.

In the case of the *plane mirror* (Fig. 156), if we consider a point source of light set up in front of it, and trace the course of the rays, which are reflected from various parts of it, by means of the general law of reflection, we shall find that they all diverge from a point which is situated behind the mirror on the perpendicular drawn from the source to the mirror, and as far behind the mirror as the source is in front of it. This point is called the *image* of the source.

We may prove this formally as follows: The ray *SM* which falls perpendicularly on the mirror at *M* will be reflected perpendicularly back, and will proceed, therefore, as if it came from some source back of the mirror on the prolongation of *SM*. The ray *SO* makes the angle α with the normal to the mirror at *O* and the reflected ray *OP* makes an equal angle with the normal. The light proceeds along *OP* as if it came from a source situated on the prolongation of *OP* behind the mirror. The prolongations of *SM* and *PO* will meet at some point *I*. This point is the image of the source. For the triangles *SOM* and *IOM* are equal, having the angles *OSM* and *OIM* each equal to α, the angles *SMO* and *IMO* right angles, and the base *OM* common: therefore the side *SM* is equal to the side *IM*. This conclusion does not depend on any particular choice of the incident ray, but will be true for all incident rays coming from the source, so that these rays will all proceed after reflection as if they came from the point *I*, which is therefore the image. The image lies on the normal to the mirror from the source, and is as far behind the mirror as the source is in front of it.

When an extended body or object is placed in front of a plane mirror, each point of it from which light can reach the mirror is matched by an image behind the mirror. The light received by the observer looking in the mirror seems to come from an object behind the mirror. Each point of the image is on the normal from the corresponding point of the source to the plane of the mirror, so that the image is of the same height as the object. The top of the object is the top of the image, so that the image is *erect*. When observing one's own image the image of the right hand appears to

be the left hand of the person seen in the mirror; the image is said to be *perverted*.

305. **The Spherical Mirror; Foci.** — To study the *spherical mirror*, we suppose a small portion of a spherical shell to be cut off by a plane section (Fig. 157). The diameter of the sphere prolonged through the middle point of the section we call the *axis*. If the reflection occurs at the inner surface of this spherical cap, the mirror is concave; if at the outer surface, it is convex. The point X where the axis cuts the mirror is called the *vertex*. The center C of the sphere, of which the mirror forms a part, is the *center* of the mirror, and the radius r of that sphere is the *radius* of the mirror.

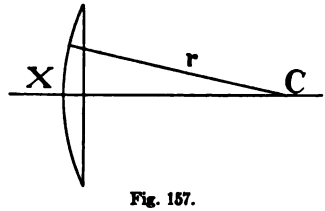


Fig. 157.

We shall consider first the case of the *concave* mirror (Fig. 158). If we suppose the source of light to be at a point S on the axis more distant from the mirror than its center, and if we trace the incident and reflected rays SM and MP from a point M on the mirror, which make equal angles with the radius drawn to that point, we find that the reflected ray cuts the axis at a point lying between the vertex and the center. We suppose that the angle between the incident ray and the axis is always small. With that limitation it may be proved that whatever be the point on the mirror at which the incident ray meets it, the reflected ray will always pass through the same point on the axis.

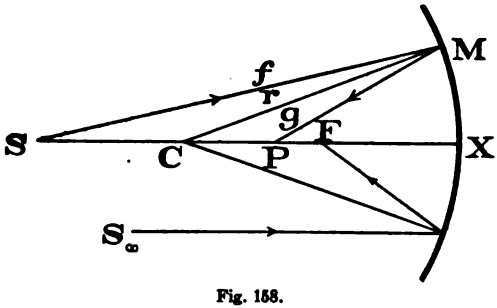


Fig. 158.

To do this we investigate the distance XP from the vertex to the point on the axis at which the reflected ray intersects it. Let us represent the distance of the source from the center by a, the distance of the point of intersection P from the center by b. Let us represent further the radius of the mirror by r, and the distances from the point M on the mirror to the source and the point of intersection respectively, by f and g. The angles SMC and PMC, being the angles of incidence and reflection, are equal, and we therefore have among the sides of the triangles the proportion $f : g = a : b$. If the mirror is small, we have, to a first approximation, $f = u$, the distance of the source from the vertex, $g = v$, the distance of the point of intersection P from the vertex. Replacing f and g

in the above proportion by u and v and using $a = u - r$, $b = r - v$, we obtain $(u + v)r = 2uv$; and dividing by uvr we get the conventional form of the formula for spherical mirrors

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}. \tag{104}$$

This equation gives the distance from the vertex of the point of intersection of the reflected ray with the axis, in terms of the distance of the source from the vertex and of the radius of the mirror; that is, in terms of quantities which are independent of the angle between the incident ray and the axis. The same point, therefore, will be a common point of intersection of all reflected rays coming from the given source. It is called the *focus* of the source.

It is plain from the construction that the source and focus may be interchanged. The two points thus related are called *conjugate foci.*

If the source is at an infinite distance from the mirror, so that $u = \infty$, the formula gives $v_\infty = r/2$. The point F thus determined lies halfway between the center and the vertex. It is called the *principal focus* of the mirror. The distance V of the principal focus from the mirror is called the *focal length* of the mirror. As the source moves up toward the mirror, or as u diminishes, v increases and the focus moves toward the center. At the center, where $u = r$, we have $v = r$ also. After the source has passed the center, the focus moves out toward infinity, and is at an infinite distance when the source is at the principal focus. During all these changes the focus is a point through which the reflected rays actually pass. It is therefore called a *real focus.*

If the source is between the principal focus and the vertex, the formula shows that the distance of the focus from the vertex is negative. This indicates that the focus stands behind the mirror. It may be shown by construction (Fig. 159) that, in this case, the reflected rays diverge from the mirror as if they came from a point behind it. This point, through which the rays do not actually pass, is called a *virtual focus.*

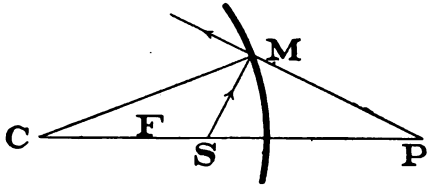


Fig. 159.

If we use x and y to represent the distances of the source S and the focus P respectively from the principal focus, we may write $u = x + V$, $v = y + V$, and substituting for u and v in the formula for the concave mirror we find

$$xy = V^2 = \frac{r^2}{4}. \tag{105}$$

The lines x and y are positive when drawn from the principal focus toward the source. They change sign together.

If the mirror is convex instead of concave (Fig. 160), a construction similar to the one already made indicates that the reflected rays from a point source always diverge from the mirror, and that the focus of any point on the axis is a virtual focus. A study of the formula shows that the principal focus lies midway between the center and the vertex, and that the focus of any real source on the axis lies between the principal focus and the vertex.

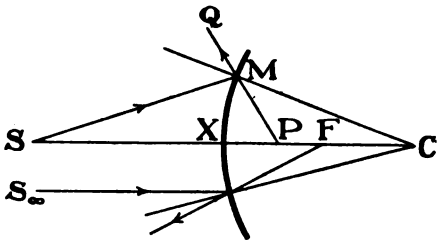


Fig. 160.

To prove that such a focus exists, we must, as before, find the position of the point of intersection P of the reflected ray MQ with the axis. We represent the radius by r, and set $SM = f$, $PM = g$, $SC = a$, $PC = b$. The exterior angle SMQ of the triangle SMP is bisected by the normal CD, and hence we have the proportion among the sides

$$f : g = a : b.$$

If the point of incidence M is so near the axis that the angle made by the ray with the axis is small, we have approximately $f = u = SX$, $g = v = PX$. Setting $a = u + r$, $b = r - v$, and using these values in the proportion, we have

$$(v - u)r = -2uv,$$

or in the conventional form

$$\frac{1}{u} - \frac{1}{v} = -\frac{2}{r}.$$

This formula gives v, the distance of the point P from the vertex, in terms of quantities which do not depend on the position of the point M or on the direction of the incident ray. This distance will therefore be the same for the points of intersection of all the reflected rays with the axis, and the common point of intersection will be a focus. As before, the value $V = \frac{r}{2}$ for $u = \infty$ fixes the position of the principal focus F of the mirror, and measures its focal length.

As in the case of the concave mirror, we may represent the distances from the principal focus to the source and focus respectively by x and y. In this case we have $x = u + V$, $y = V - v$. Substituting for u and v in the formula for the convex mirror we get as before $xy = V^2 = \frac{r^2}{4}$.

The formulas for the two classes of mirrors are the same in form, except for the signs of the terms containing v and r. If we agree to draw all lines from the mirror and to call those lines positive which are drawn in the direction opposite to that in which the light

travels, and those negative which are drawn in the other direction, we may write as the general formulas which represent the relations of source and focus for any mirror,

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{r}=\frac{1}{F}, \tag{106}$$

$$xy = V^2.$$

The first formula is generally used to find either v or r when u and r or u and v are given. We substitute for the given quantities their values affected with the appropriate signs, and the formula then gives the value of the quantity desired with its appropriate sign. When the light comes from a real point source the source is in front of the mirror and u is positive. It may happen, however, that the light falls on the mirror in a converging pencil, the apex of which is behind the mirror. This apex is then a *virtual source*, and its distance u from the mirror is negative.

306. The Spherical Mirror; Images. — If a small extended object stands transverse to the axis, a line may be drawn from each point of it through the center to the mirror, which will be an axis for that point, and the focus of the point will lie on that axis, or on its prolongation through the mirror. The assemblage of foci found in this way for the different points of the object constitutes what is called an *image* of the object. According as the foci which form it are real or virtual, the image is called a *real* or *virtual image*.

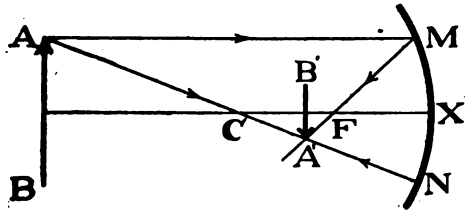


Fig. 161.

We study the position and general characteristics of the image by the aid of the following construction (Fig. 161): A ray from the point A of the object AB is drawn through the center C of the mirror. This ray will be reflected back at N along the same line, so that the focus A' of the point source A will lie on that line. Another ray from the point A is drawn parallel with the central axis CX. Since the principal focus F on that central axis is the point of intersection after reflection of all rays coming from an infinitely distant source on that axis, and therefore of all rays parallel with that axis, the ray AM will be reflected so as to pass through F. The point A' where it crosses the ray AN, or its reflected ray, will be the point where all other reflected rays from A will intersect, and will be the focus of A. Similarly the point B' will be the focus of B; and the foci of the intermediate points of the object will lie in corresponding positions between A' and B'. So long as the object lies outside the principal focus, the image is in

front of the mirror, or is real and inverted. When the object lies inside the principal focus, the image is virtual and erect.

Another way of locating the points of the image is of more general application. We draw the ray AM and its reflected ray MA' through the principal focus as before (Fig. 162). Another ray from A is drawn through the principal focus and is reflected at O. Since it passes through the principal focus it will be reflected as a ray OA' parallel to the central axis, and its point of intersection at A' with the reflected ray MA' will determine the position of the point A'. In a similar way the other points of the image are determined.

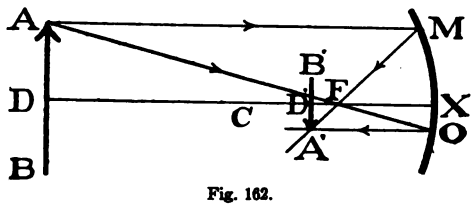


Fig. 162.

The height AD of the object and the height $A'D'$ of the image are connected by a relation which can be found from this construction. Using our previous notation, setting $FD = x$, $FD' = y$, $FX = V$, $AD = H$, $A'D' = h$, we have from the relations of the sides of the similar triangles FAD, FOX; $FA'D'$, FMX, the proportions $H : h = x : V$, and $H : h = V : y$.

Either of these formulas may be used to find h, if the required data are given. Combined they give $xy = V^2$, which we have already obtained in another way.

If we agree to draw all lines from F and to call those positive which are drawn in a direction opposite to that in which the light travels before reaching the mirror, and those negative which are drawn in the other direction, we shall have

$$H : h = x : -V, \quad \text{or} \quad h = -\frac{HV}{x}, \tag{107}$$

so that h is negative. This indicates that it is drawn in the opposite direction from the axis from that in which H is drawn, or that the image is inverted.

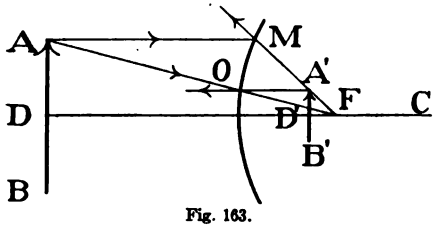


Fig. 163.

The construction for the image formed by a convex mirror is made in a similar manner (Fig. 163). With the convention just made with respect to the signs we find $H : h = x : V$, $H : h = V : y$, so that h is positive, or the image is erect. It is always virtual, since all the foci of real point sources in a convex mirror are virtual.

Real images can always be studied by throwing them upon screens. Virtual images can only be examined by the eye.

307. Mirrors of Large Aperture.— The solid angle subtended by a spherical mirror from its center is called its *aperture*. The mirrors which we have considered hitherto, and to which our former statements apply, were of very small aperture. When the aperture becomes considerable, the approximations upon which those statements depend no longer hold good. We find in such cases that no

definite focus exists. The rays reflected from the different points of the mirror cut one another in such a way as to form a specially illuminated surface. In the plane diagram this becomes a special line called the caustic.

When the source is not placed on the axis, the reflected light passes from a small portion of the mirror in such a way as to determine a line in space through which the rays pass, and which is in consequence specially illuminated. This is called the first focal line. It is at right angles to the plane of incidence. After passing through this they again determine a specially illuminated line, which is in the plane of incidence. This is called the second focal line.

EXAMPLES, XVI

1. *Two plane rectangular mirrors are set with their edges in contact so as to include an angle of θ degrees. An object is put in the angle between them. Show that the images all lie on the circumference of a circle whose center is at the intersection of the mirrors.*

In Fig. 164 the point O represents the object, the point I_1 the image formed by reflection in the left-hand mirror, the point I_2 the image of this image, formed by reflection in the right-hand mirror, etc. From the law of the position of an image formed in a plane mirror (§ 304) it follows at once that O and I_1 are at equal distances from any point on the left-hand mirror, I_1 and I_2 at equal distances from any point on the right-hand mirror, so that if the point of intersection of the mirrors is selected, the three points are at equal distances from it. The same is true of all the images, which therefore lie on a circle whose center is the point of intersection.

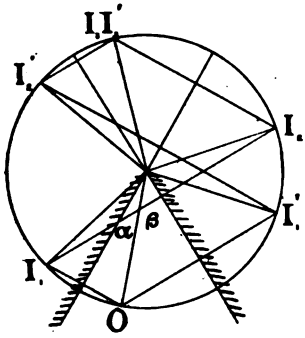


Fig. 164.

2. *To find the angles between the lines drawn from the center to the successive images.*

In Fig. 164, let α and β represent the angles between the mirrors and the line drawn from the center to the object. Then the angle subtended by the arc OI_1 is 2α, that subtended by the arc OI_1' is 2β. The arc I_1I_1' subtends the angle $2\alpha + 2\beta$, the arc I_1I_2 the angle $4\alpha + 2\beta$, so that the arc $I_1'I_2$ subtends the angle 2α. Similarly the arc I_1I_2' subtends the angle 2β. The angles are alternately 2α and 2β.

3. *To find the number of images when the angle between the mirrors is commensurable with two right angles.*

When an image falls behind both mirrors, in the angle between their planes, it cannot be reflected. Starting from O and going round the circle toward the left, the angular interval between the points determined by the object and the successive images is alternately 2α and 2β, so that the image I_3 on the left will be separated from O by the angular interval $4\alpha + 2\beta$. Similarly the image I_3'

on the right will be separated from O by the angular interval $2\alpha + 4\beta$. Now in the diagram $\alpha + \beta = 60°$, so that $3(\alpha + \beta) = 180°$, and the sum of the angular intervals determining the positions of I_1 and I_1' is $6(\alpha + \beta) = 360°$, so that these images coincide. They both lie behind the planes of both mirrors, because the angle between the line from the center to O and the plane of the left-hand mirror is $3\alpha + 2\beta$, that between the same line and the plane of the right-hand mirror is $2\alpha + 3\beta$. Counting the coincident images twice, there are 6 images, and in general if the ratio of 2π to the angle between the mirrors is n there are n images.

The propositions here considered explain the operation of the kaleidoscope.

4. *Show that light from a point source placed at the focus of a parabolic mirror will be reflected in a beam of rays parallel to the axis of the paraboloid.*

By the geometry of the parabola, any diameter and the line drawn from the focus to the point where the diameter meets the curve make equal angles with the tangent.

5. *Show that light radiating from one of the foci of a prolate ellipsoid formed by rotating an ellipse about its major axis, and reflected at the surface of the ellipsoid, will converge to the other focus.*

Reflection will be the same in any plane containing the axis. It is a known property of the ellipse that the radii vectores drawn from the foci to any point on the ellipse will make equal angles with the normal at that point. Such radii therefore fulfill the conditions of reflection and light proceeding along one of them will be reflected along the other.

All the light which leaves one focus will be reflected to the other focus.

Refraction of Light. Lenses

308. Law of Refraction. — Attention has already been called to the fact that when a ray of light passes from one medium into another, it is refracted or bent at the surface which separates the media. The ray in each medium is straight. For very many years the attempts which were made to discover any law governing this refraction were fruitless. The Greek astronomer Ptolemy (70–147) was the first to investigate this question, by the study of the refraction between air and water, air and glass, and water and glass. He could find no general law, and was forced to content himself with an empirical table, giving the angles of refraction corresponding to certain angles of incidence. Even Kepler (1604), who investigated the same question by the help of the measurements of Ptolemy and others, did not perceive the true law. It was first discovered by Snell and was first published by Descartes (1644).

To state the law of refraction, we consider a ray of light SO (Fig. 165), incident in what we may call the first medium upon a surface separating two transparent media at the point O, and making an angle α with the normal to the surface at that point. A ray OP

will then exist in the second medium which will make the angle β with the normal to the surface. The angle α is called the angle of incidence; the angle β, the angle of refraction. The law of refraction is then stated as follows:

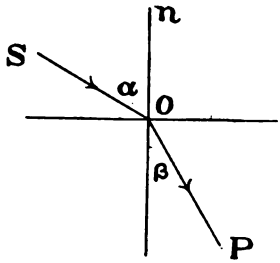


Fig. 165.

The incident and refracted rays lie in a plane which contains the normal to the refracting surface, and the ratio of the sine of the angle of incidence to the sine of the angle of refraction is constant. In terms of the symbols which we have adopted we express the law of refraction in the formula

$$\frac{\sin\alpha}{\sin\beta} = n, \tag{108}$$

in which n is a constant.

The numerical value of this constant depends upon the nature of the media which lie on either side of the refracting surface. When the media are specified, both as to their nature and as to the order in which they are considered, the constant is called the *index of refraction* from the first to the second. Thus, if we consider the incident ray in water, refracted at a surface where the water meets a block of glass, we call the ratio of the sines of the angles of incidence and refraction in this case the index of refraction from water to glass. If only one medium is specified, it is assumed that it is the second medium, in which is the refracted ray, and it is assumed that the first medium is either air, or better, vacuum. The constant in this case is called the *index of refraction* of the second medium. For example, when light passes from air into water, the ratio of the sines is 1.333. This number is called the index of refraction of water.

Experiment shows that, for all substances with which we are at present concerned, the index of refraction is greater than unity, so that the angle of refraction is less than the angle of incidence. If the index of refraction between two media is greater than unity, the second medium, in which is the refracted ray, is said to be optically denser than the other. The term is a convenient one, if we are careful not to take it to mean the actual density of the medium.

If the refracted ray is turned back on itself by reflection in a plane mirror, or if a source of light is placed in the line of the refracted ray, the ray which proceeds through the second medium

will be refracted at the original point of incidence, and the refracted ray thus obtained will coincide with the original incident ray. The index of refraction between two media when they are taken in one order is therefore the reciprocal of the index of refraction when they are taken in the reverse order. This may be represented in symbols by the formula $n_{12} = \frac{1}{n_{21}}$, in which the order of the subscripts represents the order in which the media are taken in determining the index of refraction.

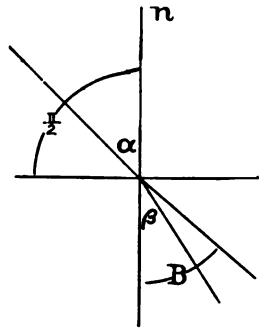


Fig. 166.

Let us consider the simple case of refraction between air and water, which is a typical one (Fig. 166). We first consider the light passing from air to water. When the incidence is perpendicular, the direction of the refracted ray is the same as that of the incident ray. As the angle of incidence α, in the air, increases, the angle of refraction β, in the water, increases also, though not so rapidly. The ratio of the sines of the two angles is always the same number. When the angle of incidence becomes a right angle, which is as large as it can be, the angle of refraction B is less than a right angle. In this case, since $\sin \pi/2 = 1$, we have $\frac{\sin \pi/2}{\sin B} = n$, or $\sin B = \frac{1}{n}$; the sine of the angle of refraction equals the reciprocal of the index of refraction. Let us next consider the light passing from water to air. When the incident ray is perpendicular, the direction of the refracted ray is unchanged. As the angle of incidence β, in the water, increases, the angle of refraction α, in the air, also increases, but more rapidly. The ratio of the sines of the two angles is the reciprocal of the index of refraction of water. When the angle of incidence B is so great that its sine is equal to the reciprocal of the index of refraction, the sine of the angle of refraction is equal to unity, and the angle of refraction is a right angle. The refracted ray, in this case, just emerges from the water into the air. If the angle of incidence is made still greater, it follows from the formula that the sine of the angle of refraction is greater than unity. This impossible result indicates a failure of the law of refraction. In fact, after the angle of incidence has passed the limiting value, for which its sine is equal to the reciprocal of the index of refraction, light no longer emerges into the air. This limiting value B of the angle of incidence is called the *critical angle*. Light incident at an angle which is greater than the critical angle is totally reflected within the water.

309. Fermat's Law of Least Time. — A general law was announced by Fermat, which governs both reflection and refraction. It may be stated by saying, that the time taken by light to pass from one point to another by way of a reflecting or refracting surface is a minimum. To illustrate this for the case of reflection, let

us consider (Fig. 167) two points S and P on the same side of a reflecting surface, from one of which the incident ray starts. The other point is the one through which the reflected ray passes. If we select any point M on the reflecting surface, and draw lines from it to S and P, it may be shown that the distance $SM + MP$ is a minimum when the point on the surface is so placed that the lines SM and MP lie in a plane containing the normal to the surface and make equal angles with that normal. Thus, if the point O is so placed that SO, OP, and the normal ON are in the same plane, and that the angles α and α' are equal, then the distance $SO + OP$ is the minimum value of the distance $SM + MP$, when M is any point on the surface. The lines whose lengths fulfill the condition that their sum is a minimum thus conform to the law of reflection, and represent the incident and reflected rays which pass through the two points. If light travels in a particular medium with a definite velocity, the time required for it to pass from the one point to the other by the lines which fulfill the minimum conditions is the least possible.

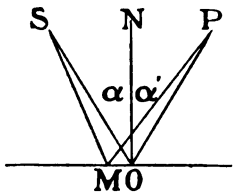


Fig. 167.

In the case of refraction Fermat assumed that the rate at which light travels is different in different media, and that the ratio of the velocities in any two media is equal to the index of refraction between those media, or that $\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}$, where v_1 and v_2 represent the velocities of light in the first and second medium respectively. If we select two points, one to serve as the source from which the incident ray comes, the other to serve as the point through which the refracted ray passes, it may be shown that this law of the velocity leads to the conclusion that the path along which light will pass from the one point to the other in the least time is that which conforms to the law of refraction.

This general principle is known as the *principle of least time*. As announced by Fermat, it was simply a hypothesis, for which no experimental proof or even theoretical argument could be given. It was subsequently shown to be a consequence of the wave theory of light, and may now be used with confidence in the solution of problems.

310. **Prisms.** — A block of any transparent substance, enclosed between two planes which meet at an edge, is called a *prism*. The angle between the planes is called the refracting angle of the prism.

We shall consider only the simple case in which the substance of the prism is glass, and the surrounding medium air. If a ray of light is incident obliquely (Fig. 168) upon one face of this prism at O it is refracted in the glass toward the normal to the first surface, and travels on in the glass until it meets the second surface at R. There it is again refracted, this time away from the normal to the second surface. The result of these two refractions, at least for many cases of incidence, is that the two refractions combine to make the emerging refracted ray deviate from the original direction of the incident ray toward the base of the prism. The angle δ between the direction of the rays SO and RP is the angle of deviation. Of course, if the original incidence is such that the refracted ray in the glass meets the second surface so that its angle of incidence there is greater than the critical angle, the ray will not emerge from the prism. Analysis shows that the deviation of the emergent ray from the direction of the incident ray is least when the refracted ray in the glass makes equal angles with the two faces of the prism. When the prism is so placed that this condition is satisfied, it is said to be in the position of *minimum deviation*. As a prism is usually a block whose cross section is a triangle, with its base perpendicular to the line bisecting the refracting angle, the condition of minimum deviation is often described by saying that in it the refracted ray in the glass is parallel with the base of the prism.

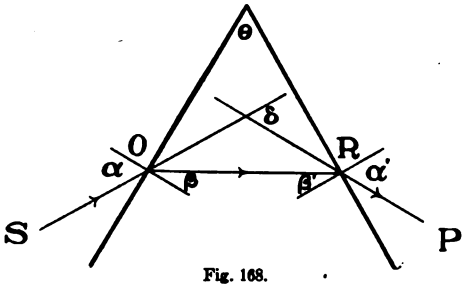


Fig. 168.

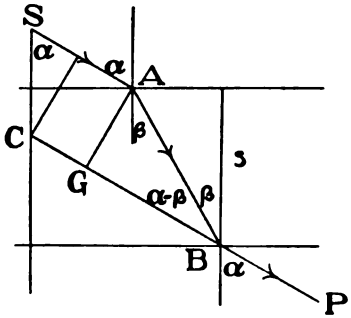


Fig. 169.

311. **Plane Parallel Plate.** — A plate of any refracting substance bounded by two plane sides, which are parallel to each other, is called a *plane parallel plate*. Light which falls upon such a plate (Fig. 169) at the angle α is refracted at the angle β and proceeds in the plate to the other face, which it meets at the angle β. On refraction at that face it emerges into the air at the angle α. The direction of the ray has not been altered, but it has been dis-

placed parallel to its original direction by an amount which depends on the index of refraction and the thickness of the plate. If the plate is so thin that its thickness can be neglected, the ray passes through the plate without deviation and with a negligible displacement.

It is interesting to find the point of intersection C of the ray BP prolonged backwards and the perpendicular SD from the source to the plate. We represent the thickness of the plate by s, and have $AB = \frac{s}{\cos\beta}$. The perpendicular AG from the point of incidence to the line BP, or the sideways displacement of the ray, is given by $AB\sin(\alpha-\beta) = \frac{s\sin(\alpha-\beta)}{\cos\beta}$ and $SC = \frac{s\sin(\alpha-\beta)}{\sin\alpha\cos\beta} = s\left(1-\frac{\tan\beta}{\tan\alpha}\right)$. If the angle of incidence is so small that its sine may be substituted for its tangent, $\frac{\tan\beta}{\tan\alpha} = \frac{\sin\beta}{\sin\alpha} = \frac{1}{n}$; and in this case $SC = s\left(\frac{n-1}{n}\right)$. The point C is in this case determined by a quantity which does not depend on the particular angle of incidence, and will be the same for all rays from S for which the angle of incidence is small. The cone of rays which leave S and fall upon the plate near the foot of the perpendicular will emerge from the plate as a cone of rays diverging from the point C, which is therefore a virtual focus. An object viewed through a thick plate will appear nearer than it really is by the distance $SC = s\left(\frac{n-1}{n}\right)$.

312. Refraction at a Spherical Surface. — If light from a point source in air is incident upon a small portion of a spherical surface bounding another medium, so placed that one of the rays or its prolongation into the medium passes through the center of the sphere, then the rays which meet the surface near the point where this central or axial ray meets it, and are there refracted, will be so deviated that they will pass through a common point. This point will be the focus of the source.

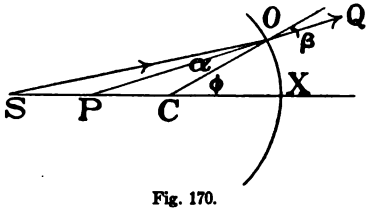


Fig. 170.

a. Concave Surface. — To prove this, let us consider a spherical surface (Fig. 170) bounding a medium, like glass, on the right from air on the left.

For convenience in dealing with the signs we shall suppose the surface concave toward the air, so that its center C lies in the air. A ray SO from the source S, on the axis SX, meets the surface at the point O, making the angle of incidence α with the radius CO, and is there refracted, the refracted ray OQ making the angle of refraction β with the radius CO produced. We denote the angle OCX by ϕ, and set $SO = f$, $PO = f'$, $SC = a$, $PC = b$, $CO = r$. Since the sides of triangles are to each other as the sines of the angles opposite, we have the proportions

$$f : a = \sin\phi : \sin\alpha$$
$$f' : b = \sin\phi : \sin\beta,$$

and from them obtain $\frac{af'}{bf} = \frac{\sin\alpha}{\sin\beta} = n$, the index of refraction. Now if the point O is near the point X, where the axis cuts the surface, we have approximately $f = u$, the distance of the source from the point X, $f' = u'$, the distance of the point P from the point X. Setting $u = a + r$, $u' = b + r$, and using these equations to eliminate a and b, we have $uu' - u'r = nuu' - nur$,

or
$$\frac{n}{u'} - \frac{1}{u} = \frac{n-1}{r}.$$

The distance u' of the point P from X is therefore determined by magnitudes which do not depend upon the particular position of the point of incidence O, provided it is sufficiently near X to permit the approximation which has been adopted. The point P will therefore be a common point through which all rays like OQ, which are produced by the refraction of rays from the source, will pass if produced backward. It is therefore a focus. Being a point through which the refracted rays do not actually pass, it is a virtual focus.

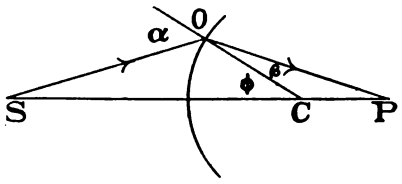


Fig. 171.

b. Convex Surface. — If we study the case of refraction at a spherical surface which is convex toward the air (Fig. 171), by using exactly the same method as before, we shall find that there is a focus, and that its position is determined by the formula

$$\frac{n}{u'} + \frac{1}{u} = \frac{n-1}{r}.$$

c. General Formula. If we compare this formula with the one obtained for the other case we shall see that they differ only in that the lengths u' and r, which are drawn in opposite directions from the surface in the two cases, appear with different signs. If we agree to call those lines positive which are drawn from the surface in the direction opposite to that in which the light travels, and those negative which are drawn in the other direction, we may write the formula

$$\frac{n}{u'} - \frac{1}{u} = \frac{n-1}{r} \tag{109}$$

as the general formula for refraction at a spherical surface. If the source is a real point source, u is itself positive. If the surface is concave toward the air, r is positive; if convex, negative. In either case the value and sign of u' will depend on the values and signs of u and r. In case u and r are both positive, u' is positive always. In case u is positive, and r negative, u' is negative, as in the illustration (Fig. 171), so long as $(n-1)/r$ is numerically greater than $1/u$. When $(n-1)/r = 1/u$, u' becomes infinite; and if $(n-1)/r$ is less than $1/u$, u' becomes positive, and the focus, which in the other case is real, becomes virtual.

If the light which falls on the surface does not proceed from a point source, but is a convergent pencil, converging to a point behind the surface, that point is considered a virtual source, and the distance u from the surface to it is considered negative.

d. Principal Foci. — If the source is at an infinite distance, we set $u = \infty$, $u' = U'$, and have $U' = \frac{nr}{n-1}$, the *second focal length.* If $u' = \infty$, $u = U$, we have $U = -\frac{r}{n-1}$, the *first focal length.* The first and second *principal foci* are located by measuring off these distances U and U' in the appropriate directions from the surface. The relation of the focal lengths to each other is given by the formula $U' = -nU$. Using the letters U and U' to designate the principal foci corresponding to the focal lengths U and U', the

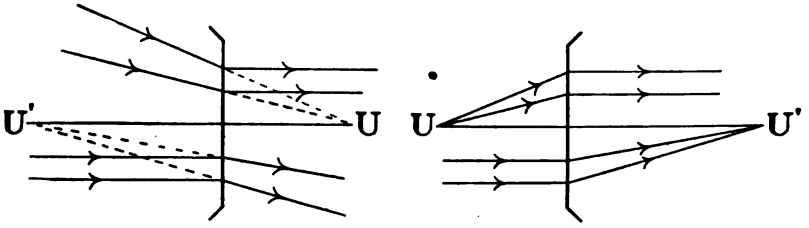


Fig. 172.

cases for positive and negative values of r are represented in Fig. 172. In these diagrammatic representations we use straight vertical lines to represent the portion of the mirror near the vertex, and indicate by the lines at the ends whether the mirror is concave or convex toward the source.

The formula 109 can be expressed in terms of the focal lengths. It takes the form

$$\frac{n}{u'} - \frac{1}{u} = -\frac{1}{U} = \frac{n}{U'}. \qquad (110)$$

313. Images. — If a small object stands transverse to the axis at the distance u from a spherical refracting surface, it will have an image corresponding to it standing at the distance u' from the surface. The properties of the image can be found by construction (Figs. 173a, 173b). A ray AM from a point A of the object, so drawn that it, or its prolongation, will pass through U, will proceed after it meets the surface parallel to the axis. Another ray AN parallel with the axis will proceed after it meets the surface so that the line representing it, or its prolongation backwards, will pass through U'. The intersection A' of the two refracted rays will be

the point of the image corresponding to A. Similarly the point B' of the image, corresponding to B, and all the intermediate points may be constructed.

Since a line AC drawn from A perpendicular to the surface, and so through its center C, will be a line along which a ray will proceed

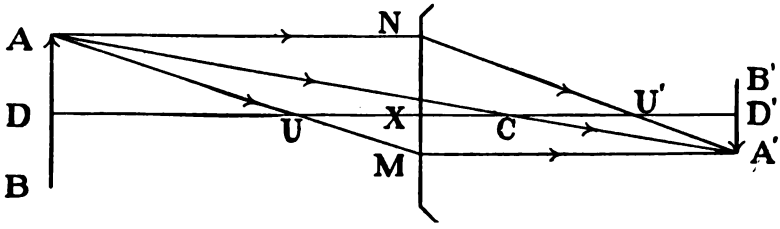


Fig. 173a.

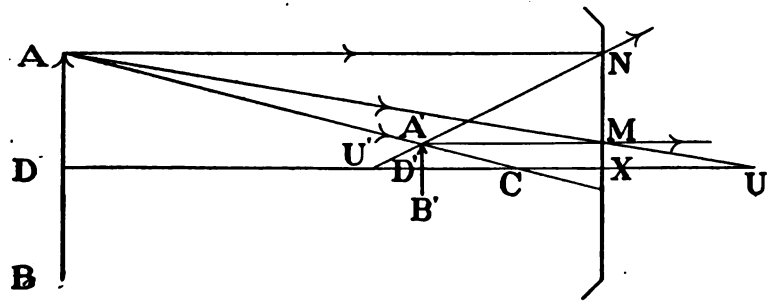


Fig. 173b.

without bending by refraction, it will also pass through the image point A', and may be used with either of the refracted rays already described to find the image point.

By the help of these constructions we may obtain useful formulas connecting the heights H and H' of the object and its image. To find one of these we notice first that the portion of the surface which is actually used to form the image is so small that we may consider a line in it, such as XM or XN, to be straight and perpendicular to the axis. Setting $UD = x$, $U'D' = x'$, $AD = XN = H$, $A'D' = XM = H'$, we have from similar triangles,

$$x : U = H : H'$$
$$U' : x' = H : H'.$$

If the other quantities are known, each of these proportions gives H'. To determine the sign of H' we have to adopt a convention with regard to the signs of x, x', U, U'. It is most convenient to consider them positive if they are drawn from the principal foci in the direction in which the light travels. With this convention, in the case of the concave surface, x and U are both negative, x' and U' both positive, and H' is of the same sign as H, so that the image is erect. In

the case of the convex surface, x is negative and U positive, x' is positive and U' negative, and H' is of the opposite sign from H, so that the image is inverted.

By combining the two proportions we get

$$xx' = UU' = -nU^2, \tag{111}$$

so that x and x' are always of opposite sign.

314. Refraction out of Medium. — We now turn to the consideration of the case in which light proceeding from a point source within some refracting medium falls on a spherical surface which separates it from air. The proof that such a source will have a focus, leading to a formula connecting the positions of the conjugate foci, will be essentially the same as before.

We designate by v' the distance of the source from the surface, positive if it is actually in the medium, negative if it is a virtual source formed by converging rays in the medium, by v the distance of the focus from the surface, by s the radius of the surface, and by n' the index of refraction from the medium into air. The formula connecting the distances of the conjugate foci is then

$$\frac{n'}{v} - \frac{1}{v'} = \frac{n' - 1}{s}.$$

The index of refraction n from air into the medium is the reciprocal of n'; and substituting for n' we get

$$\frac{n}{v'} - \frac{1}{v} = \frac{n - 1}{s}. \tag{112}$$

As in the other case, when $v = \infty$, we have $V' = \frac{ns}{n-1}$; when $v' = \infty$, $V = -\frac{s}{n-1}$, and $V' = -nV$. Using these focal lengths Equation 112 becomes

$$\frac{n}{v'} - \frac{1}{v} = -\frac{1}{V} = \frac{n}{V'}. \tag{113}$$

The diagrams for the previous case apply to this one also if we reverse them or consider the light traveling from the right instead of from the left.

Calling y' and y the distances of the source and the focus from the principal foci V' and V, we obtain as before for the ratio of the heights h' and h of the object and its image the relations

$$y' : V' = h' : h,$$
$$V : y = h' : h,$$

and

$$yy' = VV' = -nV^2. \tag{114}$$

315. Lenses. — A transparent body which is bounded by two spherical surfaces is called a *lens*. A line drawn through the lens, perpendicular to both its bounding surfaces, may be called its *axis*. If the source of light is placed on the axis, refraction will occur at the first surface and the refracted rays will proceed as if they were coming from or going toward the focus formed by this surface. When they meet the second surface, they will again be refracted,

and their new directions will determine a second focus. This focus is the focus of the source formed by the lens. Its position may be found, as our description has indicated, by using twice the formula for the focus due to a single spherical surface.

316. **Convex Meniscus.** — The lens which can be studied most satisfactorily as the typical lens is called the *convex meniscus*. It

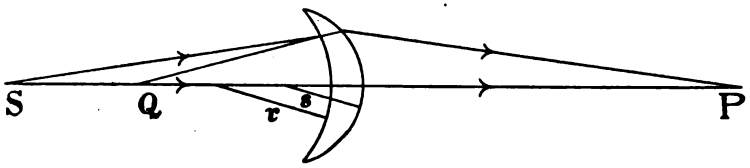


Fig. 174.

is bounded by two spherical surfaces (Fig. 174), both concave toward the source of light, if it is real, so that the radii r and s of its faces are both positive. If we suppose it to be so thin that its thickness may be neglected in comparison with the focal lengths of its faces, we may obtain the formula connecting its conjugate foci from those already obtained for its faces. For the first face, through which the light enters the glass, we have, using our previous symbols,

$$\frac{n}{u'} - \frac{1}{u} = \frac{n-1}{r}.$$

For the second face, through which the light passes into the air, we have

$$\frac{n}{v'} - \frac{1}{v} = \frac{n-1}{s}.$$

If the lens is thin, the focus Q formed by the first face, at the distance u' from it, is the source for the second face and is at the same distance from it, so that $v' = u'$. In this case we obtain at once

$$\frac{1}{v} - \frac{1}{u} = (n-1)\left(\frac{1}{r} - \frac{1}{s}\right). \tag{115}$$

If u is the distance of the source from the lens, then v obtained from the formula is the distance of the focus from the lens. Obviously the source and focus are interchangeable. They are called *conjugate foci*, and the distances u and v are *conjugate focal distances*. We use the symbols F, F' to represent what either u or v becomes when the other becomes infinite. They are called the *focal lengths* of the lens. If $u = \infty$, or if the source is at an infinite distance, $\frac{1}{v} = \frac{1}{F'} = (n-1)\left(\frac{1}{r} - \frac{1}{s}\right)$. In the case before us, since $r > s$, F' is

negative and the *principal focus* of the light is on the farther side of the lens. If $v = \infty$, so that the focus is at an infinite distance, $\frac{1}{u} = \frac{1}{F} = -(n-1)\left(\frac{1}{r} - \frac{1}{s}\right)$, and F is positive.

The signs in this general formula are such as they should be if all distances are measured from the lens, and considered positive if they are drawn in the direction opposite to that in which the light travels. The formula holds for all sorts of lenses, that is, whatever the signs of r and s may be, and for either real or virtual sources, that is, whether u is positive or negative. The value of v obtained from the formula, when the proper signs of the other quantities are introduced, will be correct in magnitude, and its sign will tell whether it terminates at a real or a virtual focus. If its sign is positive the focus is on the same side of the lens as the source and is therefore a virtual focus. If its sign is negative, the focus lies on the other side of the lens and is a real focus.

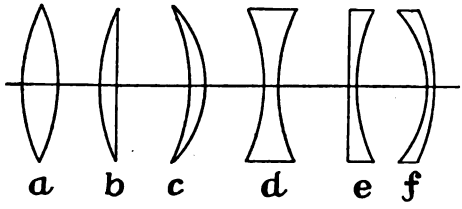


Fig. 175.

317. Converging and Diverging Lenses. — There are two general types of thin lenses, called respectively *convex* and *concave* lenses. In convex lenses the surfaces are so shaped that the lens is thickest in the middle (Figs. 175, a, b, c); in concave lenses, the lens is thinnest in the middle (Figs. 175, d, e, f).

The study of the formula $\frac{1}{F'} = (n-1)\left(\frac{1}{r} - \frac{1}{s}\right)$ shows that for the convex lenses, F' is always negative, so that light coming from an infinitely distant source converges to a real focus. For the concave lenses, F' is always positive, so that light coming from an infinitely distant source diverges from a virtual focus. In general a convex lens causes the rays which fall upon it to converge, or at least to diverge less widely than they did before they met the lens. A concave lens causes them to diverge or to converge less rapidly than they did before they met the lens. These classes of lenses are therefore called respectively *converging* and *diverging* lenses.

318. Images. — If a small object stands transverse to the axis of the lens, each point of it will act as a source, and will have its focus; the assemblage of foci thus formed constitutes the *image* of the object. The image may be found by constructions similar

to those which were used in connection with the single refracting surface. In these diagrammatic representations we use a straight vertical line to represent the portion of the lens near its center, and indicate whether the lens is convex or concave by the marks at the ends of this line. A ray from the point A of the object AB (Figs. 176a, 176b) drawn through or toward the principal focus F will proceed parallel with the axis after passing through the lens at M. Another ray from A drawn parallel to the axis and passing through the lens at N will proceed through or from the other principal focus F'.

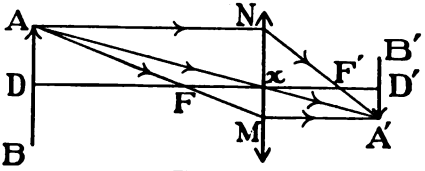


Fig. 176a.

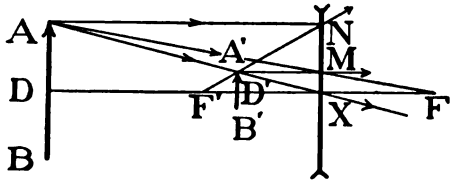


Fig. 176b.

The point of intersection of the refracted rays, produced backwards if necessary, determines the point A', which is the image of the point A. Similarly the point B' corresponding to the point B of the object, and all intermediate points corresponding to the other points of the object, can be determined.

Another construction is often convenient and sometimes indispensable. If a ray proceeds from the point A to the center of the lens, at X, it will pass through the lens without deviation. For, if the lens is thin, the two faces of the lens are parallel at the point of incidence (§ 311), and the refraction which the ray undergoes at the first surface is corrected by the refraction at the second surface, so that the ray is not deviated. Therefore the line AX will pass through the point A', and the intersection of AX with either one of the lines previously drawn through A' will determine the position of A'.

We may use this construction to determine the *focal plane*. If the object AB moves off toward infinity, the angle included by the lines drawn from the lens to its extremities A and B will grow smaller, and as the distance increases the object will become more and more like a point source, and its image will grow smaller. A very large object, however, may be placed at a distance from the lens so great that the light from each point of it has no perceptible divergence, while the angle subtended by it is still considerable. In this case the image cannot be found by the first construction, for the ray AN parallel to the axis cannot be drawn to meet the lens. We may, however, use the second construction

(Figs. 177a, 177b). The lines SX and SM represent two parallel rays coming from a point of an extremely distant source. The ray SX passes through the center at X without deviation. The ray SM passes through the focus F, and so after meeting the lens, it proceeds parallel to the axis. The point of intersection of these rays determines the focus A' of the point of the object from which the rays come. The point A' is at the same distance from the lens as the focus F', for plainly the triangles FXM, $A'XM$ are similar and equal, so that $FX = A'M$, and FX is equal to $F'X$. All the other points of an extremely distant object

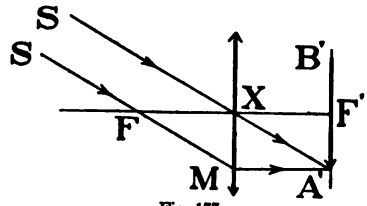


Fig. 177a.

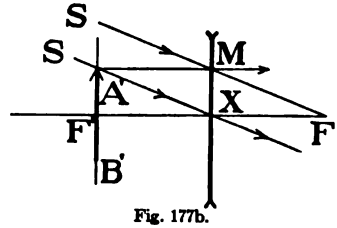


Fig. 177b.

are at the same distance from the lens, and determine a plane which stands perpendicular to the axis at the principal focus F'. This is the focal plane at F'. By symmetry there is a similar focal plane at F.

The heights H and h of the object and image are related by the proportion $H : h = u : v$, as is evident on the examination of the similar triangles ADX, $A'D'X$. If u and v are of the same sign, the image is erect; if they are of opposite signs, it is inverted (Figs. 176a, 176b).

If we transfer the origin of the lines in the diagram to the principal foci, and set $FD = x$, $F'D' = y$, we have from similar triangles

$$x : F = H : h,$$
$$F' : y = H : h.$$

If we consider those lines positive which are drawn in the direction in which the light travels, and use the appropriate signs, h may be determined in magnitude and sign. We obtain from these proportions the convenient equation

$$xy = FF' = -F^2 \qquad (116)$$

The lines x and y are therefore always of opposite sign, or are drawn in opposite directions from the principal foci F and F' respectively.

319. Thick Lenses. — When the thickness of the lens is so great that it cannot be neglected; foci and images will still be formed, but their distances from the faces of the lens will involve the thickness of the lens, and the focal lengths, measured from the faces, will not necessarily be equal. Let us return to the consideration of the typical case (§ 316) and develop the formula for the lens by taking the thickness of the lens into account.

The formulas for the two faces are

$$\frac{n}{u'} - \frac{1}{u} = \frac{n-1}{r},$$
$$\frac{n}{v'} - \frac{1}{v} = \frac{n-1}{s},$$

but in this case (Fig. 178) we cannot set $u' = v'$, as we did when we neglected the thickness of the lens. The distance v' will be greater than u' by the thickness of the lens, which we shall call, for convenience, nc. Substituting $v' = u' + nc$ and eliminating u' we obtain

$$c = \frac{1}{\frac{1}{v} + \frac{n-1}{s}} - \frac{1}{\frac{1}{u} + \frac{n-1}{r}}.$$

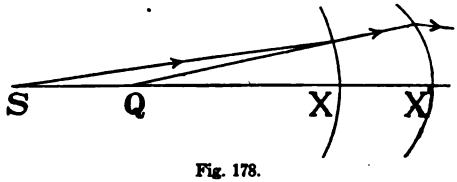


Fig. 178.

This equation enables us to find v when the other quantities are known.

If we set $u = \infty$, $v = F'$, we get

$$F' = -\frac{1}{n-1} \cdot \frac{rs + cs(n-1)}{r - s + c(n-1)}; \tag{117a}$$

and from $v = \infty$, $u = F$,

$$F = \frac{1}{n-1} \cdot \frac{rs - cr(n-1)}{r - s + c(n-1)} \tag{117b}$$

These values of F and F' determine the positions of the principal foci.

If we set as before (§§ 312, 314) $U = -\frac{r}{n-1}$, $V = -\frac{s}{n-1}$, and use $e = c - U + V$, these formulas take the forms

$$F' = \frac{V(e-V)}{e}, \qquad F = \frac{U(e+U)}{e}. \tag{118}$$

Unless r and s are equal and of opposite sign, as in the cases of the symmetrical double-convex and double-concave lenses, these focal lengths are unequal.

It was shown by Gauss (1840) that the image formed by a thick lens can be determined by a construction that is in general similar to that used for the same purpose with a thin lens. Let us set

$$\Phi = \frac{1}{n-1} \cdot \frac{rs}{r - s + c(n-1)} = -\Phi'. \tag{119}$$

Starting from the principal foci determined by F and F' we measure off Φ and Φ' in the directions determined by their signs, in the direction in which the light travels if the sign is positive. The ends of these equal lines will determine two points, called the *principal points* of the lens. The planes drawn through them perpendicular to the axis are called *principal planes*, and have the property that an object in one of them has its image in the other, erect and of the same height as the object. That is (Fig. 179), rays of light converging toward a point A on the first principal plane M, to form a real or virtual object, will be so refracted by the lens that, after emergence, they will proceed on the other side of the lens as if they came from the point A' at the same height on the second principal plane N. The two distances Φ and Φ' are the true *focal*

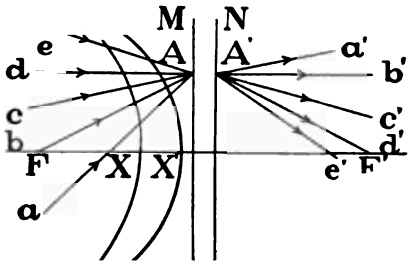


Fig. 179.

lengths of the lens. As may be seen from the formulas, the distances of the principal planes from the points X and X' respectively are

$$\left.\begin{aligned} XM = \Phi - F &= \frac{cr}{r - s + c(n-1)} \\ X'N = \Phi' - F' &= \frac{cs}{r - s + c(n-1)} \end{aligned}\right\} \tag{120}$$

taken positive when measured from X in the direction in which the light travels.

By the aid of these principal foci and principal planes the images of objects may be constructed. For example (Fig. 180), from the point A of the object draw a line through the focus at Φ to meet the first principal plane at C. It will have a ray corresponding to it on the other side of the diagram which will leave the second principal plane at the point C', whose distance from the axis is the

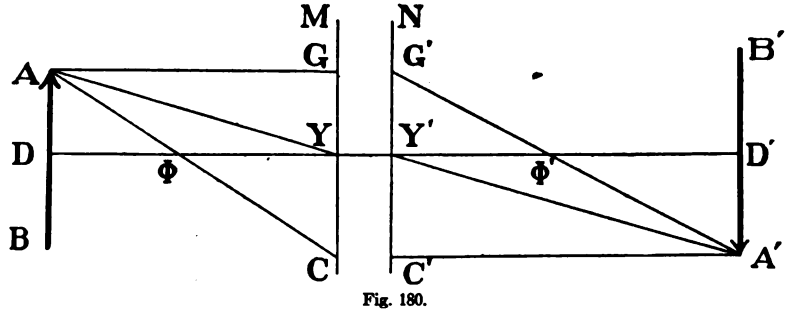


Fig. 180.

same as that of C. Since the incident ray passed through the focus, the emergent ray $C'A'$ will be parallel to the axis. Another ray from A drawn parallel to the axis, and meeting the first principal plane at G, will proceed afterwards as if it had left a point G' on the second principal plane at the same height above the axis as G. Since the ray AG is parallel to the axis, the emergent ray will pass through the principal focus Φ'. The point A', at which the two emergent rays thus drawn intersect, will be the point of the image corresponding to the point A of the object.

Setting $\Phi D = \xi$, $\Phi' D' = \eta$, and the heights of object and image H and h, we have by similar triangles

$$\xi : \Phi = H : h,$$
$$\Phi' : \eta = H : h$$

and hence

$$\xi\eta = \Phi\Phi' = -\Phi^2. \tag{121}$$

These formulas are similar to those obtained for the thin lens.

If we measure the distances ϕ and ϕ' from the object and image to the first and second principal planes respectively, we may set $\phi = \Phi - \xi$, $\phi' = \Phi' - \eta$, and substituting in the last equation for ξ and η we get

$$\frac{1}{\phi'} - \frac{1}{\phi} = \frac{1}{\Phi'} = -\frac{1}{\Phi}. \tag{122}$$

Again, substituting in the proportions, we have

$$\frac{H}{h} = \frac{\Phi - \phi}{\Phi} = \frac{\Phi'}{\Phi' - \phi'} = \frac{\Phi + \Phi' - \phi}{\Phi + \Phi' - \phi'} = \frac{\phi}{\phi'},$$

so that the heights of object and image are proportional to their distances from the principal planes.

The ray AY to the point Y where the axis cuts the first principal plane leaves the second principal plane at Y' and proceeds to A'. The lines AY and $A'Y'$ are parallel, for from what has just been proved the triangles $AYD, A'Y'D'$ are similar, and therefore the rays AY, $A'Y'$ make equal angles with the axis. This property can sometimes be used to advantage in the construction of images.

We can show as in § 318 that the various points of an extremely or infinitely distant object will have their foci on a plane standing perpendicular to the axis at the principal focus. This is the *focal plane* of the lens. By using it we may construct the emergent ray corresponding to any given incident ray. Let AC (Fig. 181) represent any incident ray meeting the first principal plane at C. It will proceed after emergence as if it passed through the point C' on the second principal plane. Draw a ray parallel to it through the principal focus Φ, meeting the first principal plane at D. This ray after emergence will proceed from D' parallel with the axis, and will meet the focal plane through Φ' at P. The other ray AC, parallel to ΦD, will meet the focal plane at the same point, so that the line $C'P$, drawn from the point C' where the emergent ray corresponding to AC leaves the second principal plane to the point P where it meets the focal plane will represent the emergent ray.

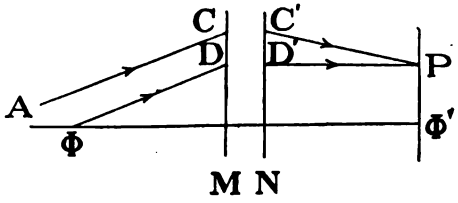


Fig. 181.

320. Lenses of Large Aperture. — The formulas which have been given and the statements which have been made with respect to lenses, hold true when only a small portion of the lens around its axis is used, and then as a first approximation. When lenses of large aperture are used, the light from a source on the axis does not come to an exact focus. The intersection of the various rays determines an especially illuminated surface, called the caustic surface.

Furthermore, in the case of such lenses, the dimensions of the image are not in the same proportion to each other as those of the object. These defects of the image are said to be due to the *spherical aberration* of the lens. By combining two lenses, so shaped that the spherical aberration of one is in the opposite sense to that of the other, a single object glass or eyepiece can be constructed which is free from spherical aberration.

EXAMPLES, XVII

1. *To find the optical center of a lens.*

The optical center is a point on the axis of the lens so placed that all the rays which pass through it after refraction at the first surface emerge at the second surface in directions parallel to their original directions.

From the centers R and S of the two faces (Figs. 182 a, b) draw parallel radii RO and SP. The tangent planes at the points O and P will also be parallel, and light which enters at O and emerges at P is equally refracted at the two surfaces, and emerges parallel with its original direction. By construction, the triangles

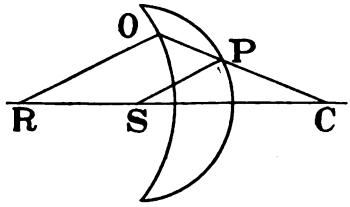


Fig. 182a.

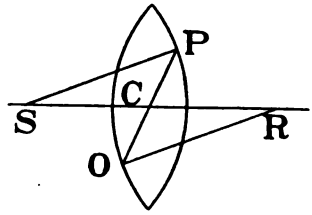


Fig. 182b.

ROC, SPC are similar, and therefore $RC : SC = RO : SP$. But $RO = r$, $SP = s$, for every position of the points O and P, and therefore the ratio $RC : SC$ is constant, so that C is a fixed point. It is the point called the optical center.

2. *To find a formula for the position of the optical center.*

We represent the distances of C (Fig. 182) from the first and second surfaces respectively by f and g. When we pay attention to the signs, we have

$$RC = r - f, \quad SC = s - g,$$

and from the proportion of Example 1 we have

$$\frac{r-f}{s-g} = \frac{r}{s} = \frac{f}{g}.$$

Further, introducing the thickness nc of the lens, we have $nc = g - f$, nc being treated as a positive constant. From these equations we get

$$f = \frac{-ncr}{r-s}, \qquad g = \frac{-ncs}{r-s}.$$

3. *To find the source of which the optical center is the image, formed by the first surface. Also, the image, formed by the second surface, of which the optical center is the source.*

We have $\dfrac{n}{u'} - \dfrac{1}{u} = \dfrac{n-1}{r}$, and $u' = f = -\dfrac{ncr}{r-s}$.

Hence $$u = -\frac{cr}{r - s + c(n-1)}.$$

Also $\dfrac{n}{v'} - \dfrac{1}{v} = \dfrac{n-1}{s}$, and $v' = g = -\dfrac{ncs}{r-s}$.

Hence $$v = -\frac{cs}{r - s + c(n-1)}.$$

These distances u and v are the distances (§ 319, Equations 120) of the principal planes of the lens from the faces of the lens. The signs are positive when the lines are measured from the faces toward the source of light. The points where the axis meets the principal planes are conjugate foci of the optical center formed by the first and second surfaces of the lens.

4. *To find the ratio of the height H of an object in the first principal plane to the height H' of its image in a plane at the optical center.*

This is most simply done by setting up a small object in the principal plane (Fig. 183) at M and drawing a line to an end point of it from the center O of the refracting surface. This line represents a ray of the object which will proceed without deviation, since it enters the refracting surface perpendicularly to it. The line OA produced will therefore pass through the corresponding point B of the image formed at the optical center. From similar triangles we then have, using the notation of the previous examples,

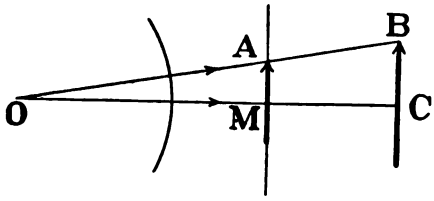


Fig. 183.

$$\frac{H}{H'} = \frac{r-u}{r-f} = \frac{r-s}{r-s+c(n-1)}.$$

Similarly the ratio of the height H' of the image at the optical center to the height h of the image of it formed by the second surface is

$$\frac{H'}{h} = \frac{s-g}{s-v} = \frac{r-s+c(n-1)}{r-s}.$$

By combination of these ratios we find $H = h$, so that the height of the object in the first principal plane is equal to the height of its image in the second principal plane. This property is used in Example 5 to determine the positions of the principal planes.

5. In the study of thick lenses it is convenient to select as a type a convex meniscus (Fig. 184) made of glass with an index of refraction n. The thickness

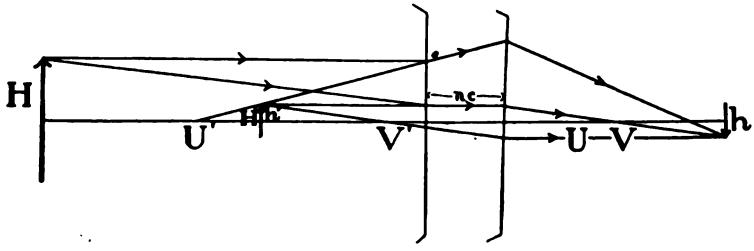


Fig. 184.

of the lens is nc, and we consider only those rays which pass through the lens in its central portion. Calling the radius of the front surface r, of the rear surface s, we have the focal lengths of the surfaces (§§ 312, d, 314) given by

$$U = -\frac{r}{n-1}, \quad U' = -nU, \quad V' = -nV, \quad V = -\frac{s}{n-1}.$$

We measure U and V from the principal foci thus determined, if positive, in the direction in which the light travels.

a. To find the principal foci.

Let us represent the distance from U' to V' by a. Then $a = nc + U' - V' = n(c - U + V) = ne$, using e to denote $c - U + V$. This distance is also equal to the sum of the distances from U' and V' to the image H' of a distant object

formed by the first surface, so that, with attention to the signs and in the notation of §§ 313, 314, we may write $a = x' - y'$.

Hence (Equations 111, 114) $a = \frac{nV^2}{y} - \frac{nU^2}{x}$.

When the object is at an infinite distance, $x = \infty$, $y = Y = \frac{nV^2}{a} = \frac{V^2}{e}$; when the image is at an infinite distance, $y = \infty$, $x = X = -\frac{nU^2}{a} = -\frac{U^2}{e}$.

The distances X and Y are measured from the principal foci U and V of the two surfaces, which are those formed by light traveling parallel to the axis in the glass. The foci determined by X and Y are the *principal foci* of the lens.

b. To find the location of an object whose image is erect and of the same height as the object.

From Example 4, this object and its image determine the principal planes of the lens.

From §§ 313, 314 we have $\frac{H}{H'} = \frac{x}{U}$; $\frac{h'}{h} = \frac{V}{y}$.

Since the image H' is the object h' of the second surface we have $H' = h'$, and hence $\frac{H}{h} = \frac{xV}{yU}$, as a general relation between the heights of the object and of the final image. If now we set $H = h$, and designate the values of x and y which are determined by that condition by $\bar{x}$ and $\bar{y}$, we find $\bar{x}V = \bar{y}U$.

Substituting for $\bar{y}$ and $\bar{x}$ respectively in the general equation for a, we find

$$\bar{x} = \frac{nU(V-U)}{a} = \frac{U(V-U)}{e}; \quad \bar{y} = \frac{nV(V-U)}{a} = \frac{V(V-U)}{e}.$$

These distances are measured from the principal foci U and V of the two surfaces.

c. To find the true focal length of the lens.

The true focal length Φ is the distance from that one of the principal foci which is determined by reference to the point U to the principal plane determined by reference to the same point. We have $\Phi = \bar{x} - X = \frac{nUV}{a} = \frac{UV}{e}$. Similarly the true focal length $\Phi' = \bar{y} - Y = -\frac{nUV}{a} = -\frac{UV}{e}$. These focal lengths are measured from the principal foci. They are equal in magnitude and opposite in sign.

d. To find the distance between the principal planes.

Using p for the distance between the principal planes, the distance from U to V may be written as $\bar{x} + p - \bar{y}$. It is also equal to $U + nc - V$.

Setting these expressions equal and substituting their values for $\bar{x}$ and $\bar{y}$, we get

$$p = nc + \frac{c(U-V)}{e}.$$

The principal planes coincide when the lens is infinitely thin, so that $c = 0$, and also when $U - V = \frac{nc}{n-1}$.

6. *To find how the distance between two images on the axis depends on the distance between the objects.*

Using x and y for the distances of the objects from the first principal focus,

x' and y' for the distances of the corresponding images from the second principal focus, we have (§ 318, Equation 116) $xx' = yy' = -F^2$. Setting $y - x = d$, the distance between the objects, we find $y' - x' = F^2 \cdot \frac{y-x}{xy} = F^2 \frac{d}{x(x+d)}$, for the distance between the images. This decreases rapidly as x increases, while d is constant, so that objects at different distances are much more nearly in focus on the same plane when they are far from the lens than when they are near it.

Optical Instruments

321. The Camera. — If a convex lens is placed in an aperture made in the wall of a chamber or camera, images of external objects will be brought to a focus, with more or less distinctness, on the opposite wall. By using a movable screen, the image of any particular object can be brought to an exact focus on it. Such a screen is used in the ordinary photographic camera. If the screen cannot be moved, as in some forms of camera, the lens is so selected as to give sharp images of objects at a certain distance. The images of objects at a less distance are not sharp, and clear pictures of them cannot be obtained. The images of more distant objects are also not sharp, but the indistinctness in them is much less marked. It is plain that, if we had at our disposal a number of lenses of different curvatures, or if we were able to modify the curvature of our lens, we might obtain a distinct image of any object, even on a fixed screen.

322. The Eye. — The eye (Fig. 185) is essentially a camera of this sort. The light enters the front of the eye through a transparent shell, called the cornea. In the normal eye the surface of the cornea is spherical. After refraction at the cornea, the light traverses a liquid called the aqueous humor, and falls on a small thick lens called the crystalline lens, by which it is again refracted. Emerging from the lens it traverses another liquid called the vitreous humor, and comes to a focus on the retina. The optic nerves are distributed over the retina, and are affected by the light which falls upon them, so that the sensation of vision is produced. An excess of light is prevented by the iris, an opaque screen with a central aperture, the pupil, which varies in size in such a way as to admit only so much light as the eye can bear.

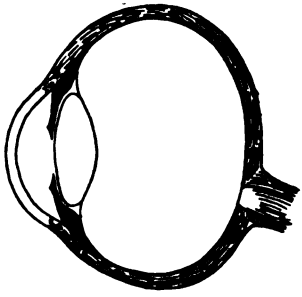


Fig. 185.

The indices of refraction of the cornea, of the aqueous humor, and of the vitreous humor, are so nearly the same that they may be considered equal. The principal features of the eye as an optical instrument may therefore be obtained by treating it as a refracting body with a spherical surface, convex toward the light, in which a thick lens of a different index of refraction is embedded. For a normal eye the radius of the cornea is about -7.8 mm. and the common index of refraction is 1.3365. By the formula of § 312d we find that the light which enters the eye from an infinite distance is so refracted by the cornea that its focus U' lies 30.98 mm. back of the cornea. The converging beam is received upon the crystalline lens. This is a double convex lens about 3.6 mm. thick; its front surface stands 3.6 mm. behind the cornea. The radii of its surfaces are $r = -10$ mm., $s = 6$ mm. Its index of refraction relative to air is about 1.4371, and relative to the liquid which surrounds it is $\frac{1.4371}{1.3365} = 1.0752$. By applying the formulas for the thick lens (§ 319) we find that the light from an infinitely distant source, after refraction at the cornea, and further refraction at the surfaces of the crystalline lens, is brought to a focus at a point 22.8 mm. behind the cornea. This distance is the depth of the eye, and the focus thus formed falls on the retina.

As the object in view approaches the eye, its image formed by refraction at the cornea recedes backward, so that when the object is 250 mm. from the eye, the distance of its image from the cornea is 34.15 mm. An image at that distance could not be brought to a focus on the retina by the crystalline lens of the shape which has been described. By the use of certain muscles which control the shape of the lens, the curvature of its surfaces is increased, the change occurring mostly in the front surface, so that the focal length of the lens is shortened and becomes such that the object in view forms a sharp focus on the retina. This adaptation of the form of the crystalline lens to light coming from objects at different distances is called accommodation.

The angle subtended at the eye by the rays coming from the extreme points of an object is called the *visual angle* subtended by that object. It is evident that the perception of small details depends upon the visual angles subtended by them, and hence that they will be better perceived the nearer they are to the eye. We cannot, however, push the examination of details to an extreme by bringing the object very near the eye; for, as the object is brought

nearer, its image tends to recede behind the retina, and it can only be kept on the retina by increasing the curvature of the lens. This cannot be done beyond certain limits, and after the object has approached the eye within a certain distance, it can no longer be seen distinctly. Even when it can be seen distinctly, the effort involved soon tires the eye. The normal nearest distance at which objects can be seen distinctly without perceptible effort is called the *distance of distinct vision.* This distance, for the normal eye, is generally taken to be about 10 inches, or 25 centimeters.

The nearsighted eye is one in which the radius of the cornea is so small that the image formed by it of distant objects cannot be brought to a focus on the retina, even when the focal length of the crystalline lens is as long as it can be made. The images of very near objects may be brought to a focus on the retina; those of more distant objects fall in front of it. This defect is corrected by placing before the eye a concave lens, which slightly increases the divergence of the light coming from distant objects, and so throws back their images on the retina. The farsighted eye is one in which the radius of the cornea is so great that the images formed by it of near objects cannot be brought to a focus on the retina, even when the focal length of the crystalline lens is as short as it can be made. The image of distant objects may be brought to a focus on the retina; those of nearer objects fall behind it. This defect is corrected by placing before the eye a convex lens, which lessens the divergence of the light coming from near objects, so that distinct images of them are cast on the retina.

323. Optical Instruments. — By viewing an object through a lens or a combination of lenses, its visual angle may be increased, while at the same time the light coming from it can be properly brought to a focus on the retina of the eye. Such combinations are called *optical instruments.* In general they may be said to be used for magnifying objects. The ratio of the visual angle of the object when seen through the instrument to its visual angle when seen with the unaided eye may be taken as a measure of the magnifying power of the instrument.

In general the operation of any optical instrument must be studied by considering, first, the effect which the instrument will have on the cone of rays proceeding from a point of the object. These rays must be so deviated that they enter the eye either as a pencil of parallel rays or as a pencil of slightly divergent rays, coming from a virtual focus at the distance of distinct vision. If the

optical system is so arranged that the rays from any point of the object enter the eye so that the focus can be seen, we then consider the visual angle subtended by the rays entering from two extreme points of the object so as to determine the magnifying power.

In our discussion of optical instruments we shall suppose that the lenses of which they are formed are so thin that their thickness may be neglected. When this is not the case, the lens may be replaced in the diagram and description by the system of principal foci and principal planes described in the section dealing with thick lenses.

324. The Magnifying Glass. — The convex lens is used as a *magnifying glass*. If the object is small and we wish to examine its details, we may for that purpose bring it very near the eye, and so increase its visual angle. But in that position the rays from any point of it diverge so widely that they cannot be brought to a focus on the retina. By interposing a convex lens (Fig. 186) so placed that the object is a little within its principal focus the rays from the object will be rendered less divergent by passage through the lens, so that an eye which receives them after they have passed the lens can bring them to a focus on the retina. These rays appear to the eye to come from the virtual image of the object. By properly adjusting the distance between the lens and the object, this virtual image may be brought to the position of distinct vision. When examined by the eye it appears erect and larger than the object. The amount of magnification depends on the shape of the lens. It increases as the focal length of the lens decreases.

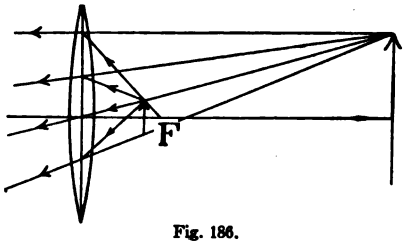


Fig. 186.

325. The Galilean Telescope. — The *telescope*, in its first form, was constructed by a Dutch optician, Jansen, who was led to it by an accidental observation made by his children. In the same form it was invented by Galileo, and is generally known by his name. The opera glass is an example of this form of telescope (Fig. 187). It consists essentially of a convex lens, or *object glass*, and a concave lens, or *eyepiece*, carried in a tube whose

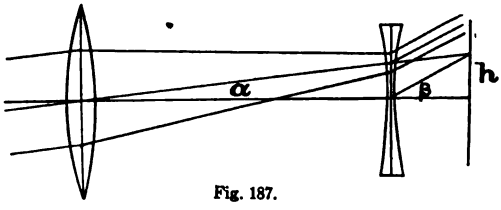


Fig. 187.

length may be adjusted. The distance between the lenses is practically equal to the difference of their focal lengths, so that the focal planes of the two lenses coincide. When the rays from a point of a distant object pass through the object glass, they are made to converge, so that the unaided eye cannot bring them to a focus on the retina. But by passage through the concave lens they are again made parallel, so that the eye looking toward that lens is able to bring them to a focus on the retina, and so sees an image of the point in the direction in which the rays enter the eye. The images of all the points of the object form the image of the object. It is virtual and erect and subtends a larger visual angle than the object.

The magnifying power of this form of telescope is given by the ratio between the focal lengths of the two lenses. For if the rays considered are those which do not deviate far from the axis of the telescope, the height h of the image formed by the object glass will be given by $h = O \tan \alpha$, if O represents the focal length of the lens and α the visual angle subtended by the object. A ray from this image also passes through the center of the eyepiece, and if E represents the focal length of the eyepiece and β the angle subtended by the image from the center of the eyepiece, we have $h = E \tan \beta$. Hence we obtain

$$\frac{O}{E} = \frac{\tan \beta}{\tan \alpha} = \frac{\beta}{\alpha},$$

if the angles are so small that the arcs may be substituted for their tangents. The ratio β/α is the magnifying power of the telescope.

326. The Astronomical Telescope. — The *astronomical telescope*, in its simplest form (Fig. 188), contains a convex lens as object glass, and a convex lens as eyepiece. The distance between the lenses is practically equal to the sum of their focal lengths, so that the focal planes of the two lenses coincide. The rays from a distant object form a real inverted image of it by passage through the object glass. After passing through the points of the image, they proceed as diverging rays. If they are then intercepted by the second convex lens or eyepiece, they will again be made parallel. On looking toward the eyepiece, the eye will see a virtual image at an infinite distance of the real image formed by the object glass. The visual angle subtended by this image is greater than that subtended by the object. Since the real image examined by the eyepiece is inverted as respects the object, the virtual image seen is

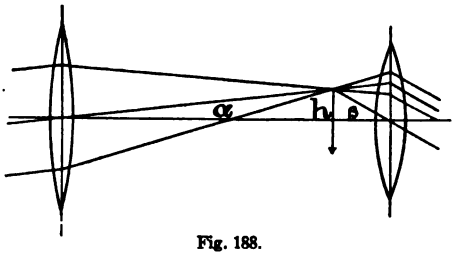


Fig. 188.

also inverted. This circumstance is of no consequence in astronomical observation. In instruments of this sort, designed for terrestrial observation, a pair of lenses is interposed between the other two, by which the real image is inverted so as to be erect. The virtual image seen is then also erect.

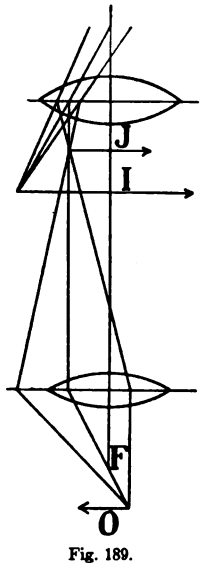


Fig. 189.

The magnifying power is given by the ratio between the focal lengths. The demonstration of this relation is essentially the same as that used in the last section.

Newton constructed astronomical telescopes in which the real image was formed by a concave mirror, and in recent years reflecting telescopes, as such instruments are called, have been constructed of large size and great efficiency.

327. The Microscope.—The *microscope* is also essentially a combination of two convex lenses (Fig. 189). The object glass is generally a lens of very short focal length. The object O to be examined is placed slightly outside its principal focus F, and the enlarged inverted image J of it, formed by the object glass, is examined with the convex lens forming the eyepiece. The adjustment is usually so made that the virtual image I viewed through the eyepiece is set at the distance of distinct vision. The magnified image in this case is inverted.

EXAMPLES, XVIII

1. *To find the principal foci and principal planes of a combination of two lenses whose principal foci and principal planes are given, as well as the distance between them.*

The method is similar to that used in the study of the thick lens. By treating a special case, and maintaining a consistent use of the signs, we obtain general formulas, which hold, when the signs are properly taken, for all types of lenses.

We consider two converging lenses; the one through which the light enters has its first principal focus at U; light leaving this focus becomes parallel on traversing the lens, and falling on the second lens, comes to its principal focus at V. We have the focal lengths $U = -U'$, $V' = -V$. The principal planes are given, so that we know the sum of the distances m and n between them; these are positive when measured from the first principal plane in each case in the direction in which the light travels; we set $t = m + n$. We have given also the distance d between the second principal plane of the first lens and the first principal plane of the second lens.

In the case represented in the diagram (Fig. 190) we have given $U = -U'$, $V' = -V$, $\overline{MM'} + \overline{NN'} = t$, $\overline{M'N} = d$.

a. If we call the distances from the foci of the first lens to the object and first

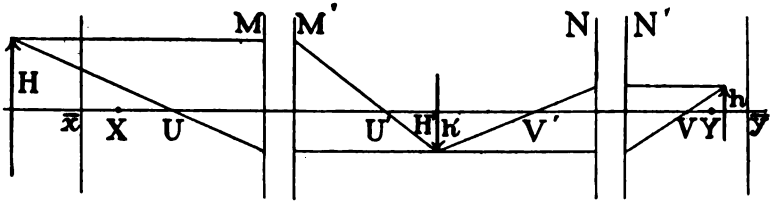


Fig. 190.

image respectively x and x', from the foci of the second lens to the first and second images respectively y' and y, we have from the similar triangles,

$$\frac{H}{H'} = \frac{x}{U} = \frac{U'}{x'}; \quad \frac{h'}{h} = \frac{y'}{V'} = \frac{V}{y};$$

and therefore $xx' = UU' = -U^2$, $yy' = VV' = -V^2$.

The distance from U' to V' is called $e = x' - y'$; and also $e = U' + d - V' = d - U + V$.

Substituting in e for x' and y' we have

$$e = \frac{V^2}{y} - \frac{U^2}{x}.$$

$$\text{If } y = \infty, \quad x = X = -\frac{U^2}{e}, \qquad \text{if } x = \infty, \quad y = Y = \frac{V^2}{e}.$$

The values of X and Y locate the principal foci of the combination. They are measured from the points U and V respectively.

b. From the proportions of 1*a*, since $H' = h'$, we have generally $\frac{H}{h} = \frac{xV}{yU}$.

When the object and final image are equal, $H = h$, and hence $\bar{x}V = \bar{y}U$, in which $\bar{x}$ and $\bar{y}$ represent the distances from the points U and V of these equal images, or of the principal planes. Substituting for $\bar{y}$ in $e = \frac{V^2}{\bar{y}} - \frac{U^2}{\bar{x}}$, we have

$$\bar{x} = \frac{U(V-U)}{e}; \text{ and similarly } \bar{y} = \frac{V(V-U)}{e}.$$

c. The focal lengths Φ and Φ' are the distances from the principal foci to the principal planes. We have $\Phi = \bar{x} - X = \frac{UV}{e}$; $\Phi' = \bar{y} - Y = -\frac{UV}{e}$.

d. The distance p between the principal planes is found by writing two expressions for the distance from U to V. We have

$$\overline{UV} = \bar{x} + p - \bar{y} = U + m - U' + e + V' + n - V = d + U - V + t.$$

Substituting their values for $\bar{x}$ and $\bar{y}$ we get for p,

$$p = t + \frac{d^2}{e}.$$

DISPERSION

328. Composition of White Light. — The rainbow naturally attracted the attention of the early students of optics. It was soon perceived that the light which reaches the eye from the bow is light from the sun, which has been reflected by the drops of rain. Descartes was able to explain the size of the bow by showing that though light will reach the eye from drops in any position after one internal reflection, yet the light that comes from those drops situated in the direction from the eye in which the bow appears to be, will be much less widely divergent, or, as he expressed it, will be denser, than the light from drops in other directions. The colors of the bow were thought to be secondary phenomena. Similar colors were observed in light reflected from drops of water, and in the light which had passed through a prism or a lens. White light was supposed to be the simplest sort of light, and the colors produced by its passage through bodies were supposed to be caused by modifications impressed by those bodies upon white light.

It was shown by Newton (1672), that this conception of white light was erroneous. Newton received a beam of sunlight, which had passed through a small hole in a window shutter, upon a prism, and cast the emergent light on a screen (Fig. 191). He found that the light on the screen, instead of being a circle, was a long colored strip. Some of the colors were more deviated than others. The colored strip he called the *spectrum*. The colors noted by him, beginning with the one which is least deviated, were red, orange, yellow, green, blue, indigo, and violet. We may interpret this experiment by saying that the index of refraction of glass for these different colors is different, and that white light, instead of being simple, contains all these colors, and is analyzed into them by the prism because of their different refrangibilities.

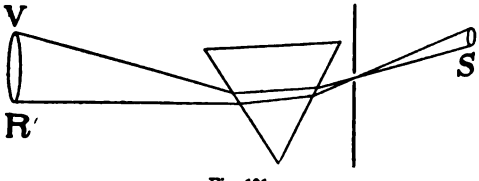


Fig. 191.

Newton proved that the refractive index was different for the different colors by placing another prism (Fig. 192) behind a small hole cut in the screen upon which a part of the spectrum was cast, and by turning the first prism so as to throw the different colors successively through the hole. When this was done, each color in turn was refracted, without any modification of color, and the deviation was different for each. To confirm this conclusion Newton superposed the colors of the spectrum by the aid of other prisms or of mirrors, and in other ways, and found that in each case white light was again obtained.

As we now use the prism, white light is passed through a narrow slit, which stands parallel to the edge of the prism, and a lens is placed in the path of the light, so that after it has passed through the prism the slit is sharply focused on the screen. Each different color which exists in white light undergoes its peculiar

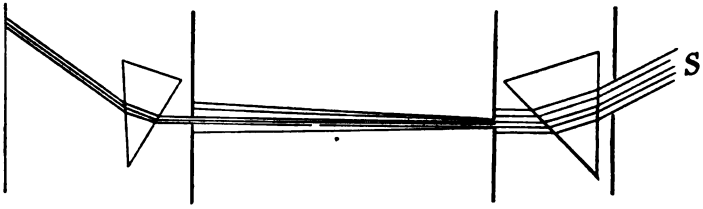


Fig. 192.

refraction in the prism, and the spectrum obtained is really a row of differently colored images of the slit. When the slit is very narrow, these images overlap very little, and the different colors are obtained with as little admixture of others as possible. Such a spectrum is called a *pure spectrum*.

The separation of white light into its constituent colors by refraction is called *dispersion*. The length of the spectrum obtained, compared with the deviation of the whole spectrum, is called the dispersive power of the prism.

329. **Chromatic Aberration.** — In view of the composite nature of white light, and the different refrangibilities of its constituent colors, it is evident that the focus formed by a lens of a point source, from which white light comes, will be different for each of the constituent colors, and hence that the image of an object viewed through a lens will be colored around the edges. This defect of the image is said to be due to *chromatic aberration*. From a few experiments which he made, Newton concluded that the dispersive power of a prism was proportional to its refractive power, so that if two prisms, or two lenses, were superposed in such a way as to annul the dispersion, they would at the same time annul the refraction. Newton therefore decided that no combination of lenses could be made which would give an image free from color. He accordingly constructed a concave spherical mirror to be used instead of the object glass in an astronomical telescope. The law of reflection being the same for all colors, the image formed by this mirror was free from color. Telescopes of this sort were much superior to the refracting telescopes then in use, and for many years quite superseded them.

330. **Achromatism.** — In 1757 Dollond, an English optician, investigated again the relation between the dispersive and refractive

powers of different kinds of glass, and found that no such proportion between them prevailed as Newton had supposed. This being so, it became possible to construct a combination of two lenses, one of flint glass and the other of crown glass, so shaped as to correct the dispersion and yet capable of refracting light and forming an image. Such a combination is called an *achromatic combination* or an achromatic lens. When an achromatic combination is used as the object glass of a telescope or a microscope, the image formed by it is almost entirely free from color at the edges, and by the use of another achromatic combination as eyepiece, the disturbing effects of dispersion are almost entirely avoided.

The dispersive power of different materials for different colors does not follow any general law, though the relative dispersive powers for the different colors are nearly the same. The achromatic combination of two lenses can therefore only completely correct for two colors. The image formed by it is still slightly colored, owing to this so-called irrationality of dispersion. By adding a third lens, a further correction can be made. In the finest microscope lenses, made after the designs of Abbe, the corrections for both spherical and chromatic aberration have been pushed so far that the lenses are practically perfect.

Velocity of Light

331. The Velocity of Light. — In our work, up to this time, we have treated the rays of light as if they marked paths along which light travels from its source. No proof has been given to justify this, nor is any needed, so long as we adhere to our original conception of light as being something which originates at external bodies and reaches the eye from them. Very different ideas prevailed among the older students of optics about the velocity with which light travels. Descartes thought that its velocity was infinitely great. Galileo, on the other hand, conceived of it as being possibly so small that it could be detected by a simple experiment.

He proposed that two observers, furnished with lanterns, should occupy two stations at a considerable distance from each other. The first observer was to expose his lantern. The second observer, seeing the light from it, was to expose his lantern in turn, and the light from it was to be observed by the observer at the first station. From the time which elapsed between the exposure of the first lantern and the reception of light from the second, Galileo hoped to determine the velocity of light. When the experiment was tried, it was found that no perceptible time elapsed between these two events, except that which was unavoidably taken by the second observer in receiving the sensation of light from the first lantern and in exposing his own. So far as this experiment went, the velocity of light was infinite.

332. Römer's Method. — It was proved not to be infinite, and its value was determined, by the Danish astronomer Römer (1676). Römer had been engaged in studying the revolutions of Jupiter's satellites. In the course of their revolutions these satellites are often eclipsed by passing into Jupiter's shadow, and the times of these eclipses can be very exactly noted. From a succession of such observations, Römer found the period of revolution of one of the satellites, and predicted the moments of its eclipses. On continuing his observations, he found that the actual times of the eclipses gradually departed from the predicted times. When the observations were analyzed, it was found that, when the earth was moving away from Jupiter, the period of the satellite appeared longer, and when the earth was moving toward Jupiter, shorter, than the mean or average value of all the observed periods. Thus, if the times of eclipse were predicted from observations made when the earth was nearest Jupiter, the observed times when the earth was farthest from Jupiter would lag behind the predicted ones. Römer concluded that these observations could be best explained by supposing that light travels with a finite, though very great, velocity. On the basis of his observations, he stated that light took 8 minutes, 18.2 seconds to travel over the mean radius of the earth's orbit.

333. Bradley's Method. — This conclusion of Römer's was for many years unconfirmed. It was not until the year 1729 that it received support from a discovery made by the English astronomer Bradley. From long-continued observations of the positions of the fixed stars, Bradley found that the stars apparently describe small paths or orbits, which are completed in one year. The paths which those stars describe which lie near the plane of the earth's orbit are short straight lines; those described by the stars which lie farther away from this plane are ellipses, which approach circles more nearly as the star lies farther from the plane. In Bradley's time it was commonly thought that light consisted of streams of minute particles, shot out in straight lines from the luminous body. On this theory of light it was easy to explain Bradley's observation. For the velocity of the light particles entering the eye would be the resultant of the velocity of light in space and of a velocity equal to that of the earth in its orbit and in the opposite direction; and the direction in which the light particle would reach the eye would be the direction of this resultant. When the earth was moving across the path of the light, the direction of this resultant would be

such that the star would appear displaced from its true position in the direction of the earth's motion. From the amount of the deviation and the known velocity of the earth in its orbit, the velocity of light could be calculated. The result obtained was in good agreement with that previously obtained by Römer's method. The phenomenon discovered by Bradley is called the *aberration* of light.

The velocity of light has since been measured by methods which do not involve astronomical observations, and the values obtained by Römer's and Bradley's methods are confirmed. While Bradley's conclusion was thus justified, the particular argument by which he reached it has long ago been discarded, because of the abandonment of the theory upon which that argument was based. The explanation of aberration by means of the theory of light now universally accepted is still under discussion. There can be no doubt that the aberration depends on the velocity of light in the way Bradley assumed it did, but it is extremely difficult to reconcile aberration with other results of experiment.

334. **Fizeau's Method.** — The first direct determination of the velocity of light, by a method which did not involve astronomical measurements, was made by Fizeau (1849). He used a toothed wheel which could be rapidly revolved at a known rate. When the wheel was still, a beam of light was sent out through one of the gaps between the teeth, was received on a mirror set up at a great distance, and reflected back through the same gap. The wheel was then set in rotation. At first, as it turned slowly, the light sent back from the mirror still passed through the same gap, but as the rotation was increased, the neighboring tooth moved forward into the path of the light, so as to intercept it. When the rotation of the wheel was doubled, the returning light passed through the next gap. In this way, by increasing the rate of rotation, the returning light was alternately intercepted and transmitted. By observing the different rates of rotation, it was easy to determine the time occupied by the light in passing from the wheel to the mirror and back again to the wheel. The value for the velocity of light determined by Fizeau in this way was about 315 million meters per second.

335. **Foucault's Method.** — Another method of determining the velocity of light was employed by Foucault (1862). He used for that purpose a small mirror, which could be rotated rapidly around an axis parallel to its face. When the mirror was at rest, light was allowed to fall on it from a slit, and was reflected by it to a fixed mirror. This mirror sent the light back over the same path, so that it fell on the slit. When the mirror was rotated, the angle

at which it received the returning light was different from that at which it had sent out the same light from the slit, and the returning light reflected from it no longer fell on the slit. The velocity of light was determined by measuring the displacement of the image of the slit sent back from the revolving mirror and the rate of rotation of the mirror. Foucault obtained a result for the velocity of light which was a little smaller than that obtained by Fizeau. This method has been used by Michelson and by Newcomb to obtain the velocity of light which is now considered the standard. Newcomb gives it as 299,860 kilometers per second. For ordinary purposes we may take it as equal to 300 million meters per second.

It was by the use of this method that Foucault proved (1850) that the velocity of light in water is less than in air.

CHAPTER XVII

PHYSICAL OPTICS

336. Physical Optics. — We have so far been able to proceed by the aid of the experimental laws of reflection and refraction. We now enter upon the examination of properties of light which cannot be described by those laws, and which eventually will lead us to the consideration of the physical theory by which all the phenomena of light are coördinated and explained. Our subject of study is called *physical optics*. Before considering the phenomena in detail, it is expedient to take a cursory view of the most important of them, as they were known to Newton and his contemporaries.

337. Colors of Thin Films. — The attention of Newton was directed to the colors seen in thin films, such as sheets of mica, or the walls of a soap bubble. These colors had been already investigated by Hooke. Newton's most important observations were made by the help of the film of air between a plane sheet of glass and a slightly convex lens laid upon it. When light was allowed to fall nearly perpendicularly upon this film, and the eye was placed above it, a set of colored rings was seen, having as a common center the point of contact between the two pieces of glass. At this point of contact there was a black spot. Newton showed that when light of one color was used, the rings were of the same color, and that the diameter of the successive rings was different for the different colors. The diameter of the ring of any order was greatest when red light was used, and least when violet light was used. When light of one color was used, the intervals between the rings were black. From this observation it was easy to explain the colors seen with white light. At any one place in the film some of the constituents of the white light will not appear, or will be extinguished, and the color actually seen is due to the mixture of the remaining constituents, which are not extinguished.

From the known curvature of the lens under which the film was formed, Newton calculated the thickness of the film at different distances from the black spot at the center of the system of rings, and found that the rings of any one color appeared at parts of the film, whose thicknesses were in the ratio of the numbers 1, 3, 5, 7, etc.

This result indicated some sort of periodicity, or regular alternation in the properties of the light.

When the film is all of one thickness, the color seen in it is the same in all parts, except for slight modifications due to the different angles of incidence upon the film of the light from the source.

338. **Diffraction.** — Other systems of colored bands of light were first observed by Grimaldi, and were also observed by Newton. They occur when light, which has passed through a very small opening, passes the edge of an obstacle, or through another small opening. They are best observed if the first opening is a narrow slit. In this case, if the straight edge of a screen is placed in the light coming through the slit and parallel with it, the shadow of this edge cast on a receiving screen is bordered with three or four narrow colored bands, standing outside the shadow. The edge of the shadow itself is not sharp, as it should be if light traveled in geometrically straight lines. The line on the screen in which it would be met by a plane passing through the slit and the edge of the obstacle is called the edge of the *geometrical shadow*. In the conditions described, the colored bands stand outside the edge of the geometrical shadow, and the illumination extends within it, gradually fading out until it becomes imperceptible.

If a narrow slit is placed in the path of the light and parallel with the first slit, the light which comes through it is a narrow band of white light, bordered on both sides by a succession of bands of variously colored light.

The phenomena here described, and others like them, are said to be due to the *diffraction* of light.

339. **Double Refraction.**—The peculiar phenomena which appear when a ray of light falls upon a crystal of calcite or Iceland spar were described by Bartholinus (1669). The light undergoes a *double refraction*. That is, it is broken by refraction into two rays, which have different directions in the crystal. If the incident ray is perpendicular to the surface, one of the refracted rays is also perpendicular to the surface. As the incident ray is differently inclined to the normal, this refracted ray follows the ordinary law of refraction. It is therefore called the *ordinary ray*. The other ray, whose law of refraction Bartholinus could not discover, is called the *extraordinary ray*. If a surface is cut on the crystal perpendicular to the axis of the crystal, the incident ray perpendicular to that surface is not doubly refracted. The axis thus determined is called the *optic axis*.

Huygens (1690) investigated the double refraction of Iceland spar, and discovered the law by which the refraction of the extraordinary ray may be described. As the description of this law depends upon the theory of light which Huygens used in obtaining it, we shall postpone the consideration of it until we discuss that theory.

Huygens discovered that other crystals possess the property of doubly refracting light, and subsequent investigations have shown that all crystals, except those of the isometric system, are doubly refracting.

340. Polarization of Light. — In his investigation of Iceland spar, Huygens discovered another remarkable property of light. To show it, we first allow a narrow beam of light to fall perpendicularly upon one face of a crystal of Iceland spar. The light is then doubly refracted, and emerges from the opposite face in two nearly parallel beams. If a second crystal of spar is placed in the path of these beams, so that they fall perpendicularly, or nearly so, upon one of its faces, each of them will ordinarily be doubly refracted, and four beams of light will emerge from the second crystal. A difference is observable between the double refraction in this case and that observed in the first crystal, for whereas the two emergent beams from the first crystal were of equal intensity, the two beams into which each of them is divided by the second crystal are, in general, not of equal intensity. Indeed, by turning the second crystal around an axis parallel with the beam of light, two positions of it may be found in which one beam of each pair disappears altogether, while the other beam of each pair has the intensity of the beam before it entered the second crystal. We may describe this phenomenon otherwise by saying that the first crystal divides the incident beam into an ordinary and an extraordinary beam of equal intensities. The second crystal divides each of these into an ordinary and an extraordinary beam, whose intensities are generally unequal. One position of the second crystal can be found for which the extraordinary beam of the first ordinary beam, and the ordinary beam of the first extraordinary beam, are suppressed or disappear. When the crystal is turned another position can be found, in which the ordinary beam of the first ordinary beam, and the extraordinary beam of the first extraordinary beam, are suppressed. The light which has been given these properties, by which it seems to differ in different directions in space transverse to the beam, is said to be *polarized*.

Huygens could give no explanation of polarization, and its exhibition by light which had passed through Iceland spar was for a long time unique. It was shown much later, as we shall subsequently consider more at length, that whenever double refraction occurs, the two rays produced are always polarized, and it was discovered also that light may be polarized by ordinary reflection and refraction, or in other ways.

Wave Theory

341. The Emission Theory of Light. — We have now passed in review the most important phenomena exhibited by light which were known to the earlier students of that subject, and which had to be considered in any attempt to construct a theory of light. From very ancient times two rival theories of light had received consideration. In one of these, light was supposed to be an actual emanation of some substance from the luminous body. In the other, light was supposed to consist in disturbances transmitted through a medium, this medium, called the *ether*, being assumed to fill all space, even those parts of it which are occupied by material bodies. In the form which this latter theory took at the hands of Huygens, these disturbances were supposed to be waves, similar in their general features to the waves of sound in air. Huygens, however, could not adapt this theory to the explanation of the transmission of light in straight lines. When the waves of the ocean enter the narrow opening of a harbor, they do not proceed in a narrow band across the waters of the harbor, leaving the water on either side of the band undisturbed, but they spread out in all directions from the entrance. In the same way, a sound which is made inside a building will spread out in all directions when it passes through an open window. This behavior of the waves with which we can actually experiment seemed to Huygens' critics to make the wave theory of light untenable, for there is nothing more conspicuous about the behavior of light than its transmission in straight lines and the fact that it does not spread out in all directions after it has passed through an opening. Largely on this account Newton refused to consider the wave theory, and set himself to develop its rival, the emission theory of light. A comparison of this theory with the wave theory, by which it was finally displaced, furnishes an excellent illustration of the methods of reasoning used in constructing a physical theory.

Newton assumes that a luminous body shoots out from itself small bodies, which he calls corpuscles. The velocity of these corpuscles is that of light. Since the corpuscles are masses, they proceed in straight lines, by the fundamental law of inertia, unless they encounter some body by which their motion is changed. By supposing them perfectly elastic, it is easy to show that when the corpuscles fall obliquely on a smooth surface, they will rebound from it in a direction which obeys the law of reflection. By supposing further that the corpuscles are sometimes strongly attracted by ordinary matter, when the distance between the corpuscle and the matter is exceedingly small, and that the corpuscles, having entered the matter, can proceed through it without interruption, Newton shows that the path of the corpuscle will be changed in direction on its entering the mass of matter in accordance with the law of refraction. As a necessary consequence of this assumption, the velocity of a corpuscle is greater in the more highly refracting medium. To account for the different colors in white light, Newton assumes that corpuscles of different mass and perhaps of other different properties correspond to them. The colors of thin films render it necessary to make an additional hypothesis which will assign to the corpuscles some sort of periodicity. The particular form of this hypothesis which Newton adopts assumes the existence of an elastic medium, called by Newton the ether, in which waves are set up whenever a corpuscle strikes the surface of a body. These waves travel along with the corpuscle, if it is reflected, and with nearly the same velocity, so that when it meets another surface the wave can be thought of as either pushing it forward to help it pass through the surface, or as pulling it backward to prevent its passing through. Newton names these two conditions the fits of easy transmission and of easy reflection. By the help of these fits the light in the thin film acquires the necessary periodicity. Another and an easier form of the same hypothesis is that given by Boscovich (1759), who assumes that the corpuscles are in continual rotation around axes perpendicular to their paths, and that their ends are different, so that as the corpuscle reaches a surface it will be reflected if one of its ends is presented to the surface, and will be transmitted if the other end is presented to the surface. In order to explain the bands produced by diffraction, Newton makes the further supposition that the corpuscles which pass near a solid edge move backwards and forwards, as he describes it, with a motion like that of an eel, and so proceed in one direction or another according to the direction in which they are moving when they escape the influence of the edge. Newton's ingenuity fails him when he attempts to apply this theory to explain double refraction, and he cannot explain polarization further than by saying that it shows that light has sides. In the subsequent development of the emission theory, which held its ground for a century and a half, additional hypotheses were made to account for the other phenomena which were discovered. It cannot escape notice that this theory, which at the outset is so simple, and which accounts so well for rectilinear transmission, and for reflection and refraction, requires a new hypothesis for each additional phenomenon which is to be explained. Each hypothesis is of service only in explaining the phenomenon which leads to its adoption, and when a new phenomenon is considered, a new hypothesis has to be made to explain it. When the consequences of any of these hypotheses are followed out, no new and previously unknown truths about light are discovered. The only necessary conclusion from the fundamental hypotheses of the theory which was an addition to the knowledge of light, turned

out to be false when it was examined by experiment. As has been stated already, Newton's principles led to the conclusion that light travels faster in optically denser bodies. This conclusion was tested experimentally by Foucault, and found to be erroneous. Light really travels slower in optically denser bodies.

342. The Wave Theory of Light. — The wave theory of light received its first impulse from Huygens (1690). He assumed that light arises at a luminous body by reason of vibrations executed by the particles of the body. These vibrations set up disturbances in a medium called the ether, and these disturbances are transmitted through the ether in accordance with the general principles of the transmission of waves. If the light is traveling in parallel rays along the x-axis, the formula which expresses the displacement y of the medium in terms of the time and the distance x from the origin is the same in form as that which was used in the study of sound; that is,

$$y = a\cos 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

When we consider any fixed point, so that x is constant, the medium at that point executes a simple harmonic motion of period T. If we consider the condition of the medium at any instant, so that t is constant, the parts of the medium lying along the x-axis will be displaced in a way represented by the curve of sines. The velocity of the wave will be given by λ/T. The amplitude a is constant.

We shall deal generally with light traveling in nearly parallel rays. In case the light is traveling from a source outward in all directions, so that the wave front is a sphere with the source at its center, the formula for the displacement in terms of the radius r of the sphere becomes

$$y = \frac{a}{r}\cos 2\pi\left(\frac{t}{T} - \frac{r}{\lambda}\right).$$

The proof of this formula need not be given here.

343. The Ether. — The existence of the *ether*, or medium in which the waves are transmitted which constitute light, was one of the necessary assumptions of the wave theory, and the extraordinary nature of the properties which must be ascribed to this medium was one of the principal reasons for the difficulty experienced by its advocates in obtaining acceptance for it. The ether must be assumed to occupy not only space which is otherwise void, but also space filled with ordinary matter. It is itself not cognizable by

the senses, and is not ordinary matter. It is only material in so far as it is associated with energy and transmits it. It is assumed to be frictionless, at least to such a degree that slowly moving bodies, like the planets, pass through it without retardation. So long as the properties by which it transmits light waves were supposed to be those of an elastic solid, the difficulty of conceiving such a body was even greater than it now is, when we suppose these properties to be those of a body in which electric and magnetic forces can arise. But the difficulty of obtaining any adequate conception of the ether is still insuperable. We can only discover, by the analysis of the experimental facts, the fundamental equations of wave transmission to which the ether conforms, and then ascribe to the ether the necessary properties to account for its behavior, even if they are not those of any known material body.

344. Huygens' Principle. — The development of the wave theory begins with a mode of conceiving the transmission of disturbances through a medium, which was first presented by Huygens, and is known as *Huygens' principle.*

Huygens perceived that if a train or succession of waves is passing through a medium, the disturbance at any point in the medium does not depend solely on a disturbance coming directly to it from the source, but is the resultant of all the disturbances which reach it at the same time from all the disturbed parts of the medium. To put it otherwise, if a point in the medium is disturbed by a wave, it thereby becomes the center of a wave disturbance, which goes out from it nearly as if it were a center of an original disturbance. If we then consider a point in the medium, and draw a spherical surface around it as center, and suppose each of the points on that surface to be in some way disturbed by a train of waves, each of those points will send a wave toward the center of the sphere, and the disturbance at the center will be the resultant effect produced by the waves which reach it, at the same instant, from the different points of the surface. In order to meet the facts of the case, especially in order to explain that a wave disturbance is not transmitted backward as well as forward, Huygens had to assume that the elementary wave sent out from a disturbed point does not transmit equal disturbances in all directions, but that the greatest disturbance is transmitted along the ray, in the direction in which the wave is advancing which disturbs the point, and that the disturbances transmitted in oblique directions diminish in intensity until, in the opposite direction to that of the ray, they fall off to nothing

(§ 212). The analysis by Kirchhoff and others of the mode of transmission of waves in a medium shows that this hypothesis of Huygens is a sound one. The principle which has been stated is known as *Huygens' principle.*

In the application of Huygens' principle we consider any set of particles in the medium transmitting the wave, which are simultaneously agitated by a luminous wave. Each of these particles becomes the center of an *elementary wave* (Fig. 193), and at any subsequent instant the disturbances which leave the centers together will have arrived at the surfaces of the spheres of equal radii drawn around those centers. According to Huygens' assumption, the envelope of the elementary waves described in this way represents the integral wave front. The normal to the wave front is the direction in which the light travels and so is the ray of geometrical optics.

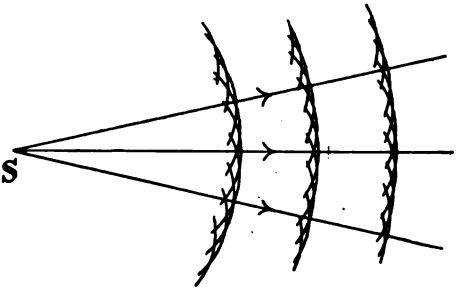


Fig. 193.

345. **Huygens' Constructions.** — Huygens explained the rectilinear transmission of light through an opening, by drawing the elementary waves (Fig. 194) around the points which stand in the line of the opening, and showing that directly beyond the opening these waves have a common envelope, while on either side of the opening they have not. He supposed the only effective part of the resulting disturbance to be that caused by this common envelope. This explanation, though ingenious, is not sufficient, and rectilinear transmission was not really explained by Huygens' form of the wave theory.

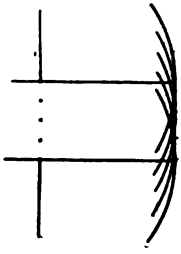


Fig. 194.

To give Huygens' explanation of reflection and refraction, we make the following construction (Fig. 195). We draw the straight line *GH* to represent the intersection of the plane of the paper with a plane reflecting surface at right angles to it. We consider a series of waves advancing obliquely toward this surface in air, forming a beam, which is bounded by the extreme rays *EA* and *FC*. At a certain instant, one of the wave fronts represented by the line *AB* will meet the surface at the point *A*. By Huygens' principle that

point will become the center of a wave disturbance, which proceeds from it in both media. We shall consider it first in the second medium. At the end of a time in which the disturbance which was at B has reached the point C, or in which the wave from A would have become a sphere in air with $AC' = BC$ as radius, the wave from A in the medium has become a sphere with A as its center but with a different radius. We shall suppose that the velocity of light in the second medium is less than that in the first, so that the radius of this sphere is less than the line BC. If we consider the light coming from any point between the points A and B, we see that it will reach the surface GH, at some point between A and C, after part of the above mentioned time has elapsed, and that there will proceed out from that point a spherical wave. When all such spherical waves are constructed, it will be found that they have a common tangent, which is a line drawn from the point C. This line is the envelope of all the elementary waves which correspond to the original disturbances in the wave front AB, and therefore represents the refracted wave. If we draw from the point A the radius AD to the point of tangency, the direction thus determined is the direction of the refracted beam. The lines BC and AD are distances passed over by the reflected and refracted waves respectively in the same time, and they are therefore proportional to the velocities v_1 and v_2 of the light in the first and second media. From the construction we see that the angles $BAC = \alpha$, $ACD = \beta$, are equal to the angles of incidence and refraction; and that $BC = AC \sin\alpha$, $AD = AC \sin\beta$, so that $\frac{BC}{AD} = \frac{v_1}{v_2} = \frac{\sin\alpha}{\sin\beta} = n$, the index of refraction. The index of refraction is therefore equal to the ratio of the velocities of light in the two media, and this ratio is presumably constant for the two media. We therefore deduce in this way the law of refraction. Huygens' hypothesis that the velocity of light is less in the denser medium, which is a necessary hypothesis of the wave theory, was confirmed by the experiments of Foucault already referred to.

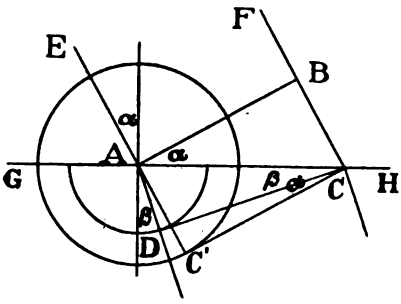


Fig. 195.

A construction that is essentially similar, in which the elementary

waves from the points on the line AC are drawn in the upper medium, will lead to the law of reflection.

By the use of the wave theory, Huygens was able to give a description of the double refraction in Iceland spar. To do so he made the supposition that the disturbance which reaches a point on the surface of the spar (Fig. 196) sets up two waves, one of which is spherical, while the other is an ellipsoid of revolution, having its axis of revolution parallel to the axis of the crystal. The first of these corresponds to the ordinary beam, the second to the extraordinary beam. By a construction which is essentially similar to the one we have used in explaining ordinary refraction, Huygens was able to determine the direction of the extraordinary ray for various incidences, and to show that, in each case, it was in complete accord with his observations.

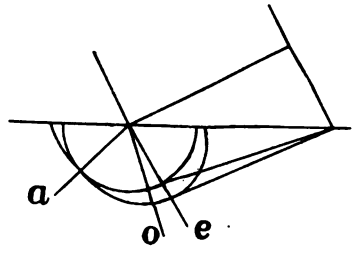


Fig. 196.

Huygens did not develop the wave theory beyond this point. In the form in which he presented it, it was really defective in that it did not explain satisfactorily rectilinear transmission. The emission theory, therefore, supported by the great authority of Newton, remained prevalent until another set of phenomena was discovered, with which the development of the wave theory really begins.

Interference and Diffraction

346. Interference of Light. — In 1803 Young discovered an effect produced by light, which of itself, and because of the consequences deduced from it, was almost conclusive in favor of the wave theory.

Young allowed a beam of light to pass through a very small opening, and received it on a distant screen. On interposing a narrow strip of card in the path of the beam, and observing its shadow on the screen, he perceived that it was not uniform, but contained a number of alternately light and dark bands, parallel with its edges. He proved that these bands were due to the combination of the two portions of the beam which had passed the two edges of the card by showing that they disappeared when the light on one side of the card was cut off by another screen.

Young subsequently described another form of this experiment, which presents its essential features very clearly. In it, light which

has passed through a very small opening is allowed to fall on two small holes or parallel slits set close together. On observing the light cast through these holes on a screen, the region in which the two beams of light overlap is seen to be crossed by a set of parallel light and dark bands standing perpendicular to the plane containing the source and the openings. When either of the openings through which the light comes is covered, these bands disappear. They cannot be obtained when the light which falls on the openings comes directly from an ordinary source. This experiment is known as *Young's experiment.*

Young explained these phenomena by the help of the wave theory. To explain the second experiment (Fig. 197), we suppose that the waves, starting from the first opening, S, which we shall hereafter call the source, set up waves at each of the two openings A and B in the screen, these waves proceeding from the openings almost as if they were independent sources. If we consider the effect at any one point P on the receiving screen produced by both these waves, we see by construction that it will depend on the relative distances AP and BP of the point from the two openings. When the difference of these distances, BC, is equal to half a wave length, or to any odd multiple of half a wave length, the two waves will reach the point in opposite phases, and by the principle of the superposition of small motions, will so counteract each other's effect that the resultant effect will be zero. The black bands that are seen correspond to this condition. At intermediate points between the black bands, for which the difference of the distances to the openings is an even number of half wave lengths, the waves from the openings will meet in similar phases. The light received at those points will be a maximum.

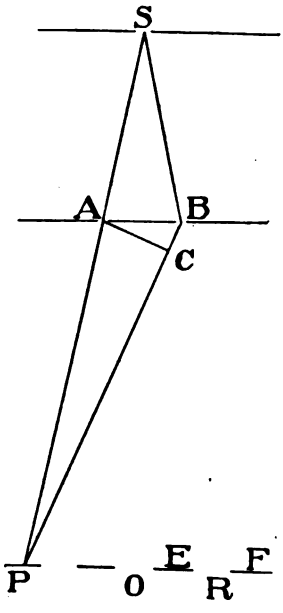


Fig. 197.

Thus, at the points O, R, etc., for which $AO - BO = 2n\lambda/2$, $n = 0$, 1, 2, . . . there will be light bands. At the intermediate points E, F, etc., for which $AE - BE = (2n - 1)\ \lambda/2$, $n = 1, 2, 3, \ldots$ there will be dark bands.

We explain the first experiment in an essentially similar way, considering that the two beams of light which combine to produce the light and dark bands are those portions of the original beam which pass close to the edges of the card and are turned into its shadow by diffraction.

The most striking single argument that was ever adduced in favor of the wave theory was that presented by the disappearance of the system of bands when the light from one of the openings was intercepted, and their reappearance when it was again allowed to pass. This effect follows as a natural consequence of the wave theory and is inconsistent with any reasonable form of the emission theory, for it seems impossible to explain how the meeting of two streams of particles will produce an effect different from that which would be produced by either stream of particles acting alone.

Waves which are superposed on one another, in the way which has been described, so as to produce variations of intensity in different parts of the field through which they are passing, are said to *interfere*, and the general phenomenon is called the *interference of light*. The waves do not interfere with one another in such a way as to destroy their individuality. Each wave passes through any point, and proceeds beyond it, as if no other wave had met it at that point. This follows as a consequence of the principle of superposition. All that is meant by interference is that the waves combine or interfere with one another at any point as respects their individual effects, so that the effect perceived is that due to their combined action.

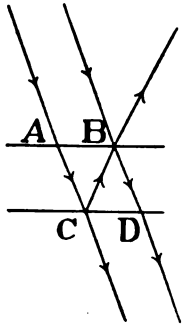


Fig. 198.

347. Colors of Thin Films. — As soon as Young had discovered interference, he found that with its help he could apply the wave theory to explain the colors of thin films. Consider a train of waves incident on the upper surface of a film (Fig. 198). Parallel and neighboring incident rays will meet the surface at the points A and B. The waves which reach A are partly reflected and partly refracted into the film. The wave in the film proceeds in the direction of the refracted ray to the point C on the second face of the film, where it will again undergo reflection within the film, and refraction into the outside medium. The reflected part will proceed again through the film, and part of it will emerge at the point B. At that point it will be superposed on a wave directly reflected there, and the effect produced in the direction of the reflected

rays will be due to the combination of these waves. If they are in the same phase they will produce an enhanced effect, so that light of the peculiar color which corresponds to the wave length will be seen. If they are in opposite phases, they will annul each other and no light will be seen. When white light is used, which we may suppose to contain waves of all periods lying between certain limits, this description shows that some of its constituents will be suppressed, while others will not be, so that the reflected light will appear colored, and the color will differ for different thicknesses of the film.

When we consider the relation between the thickness of the film and the wave length of the light which is enhanced or suppressed, it seems at first sight that the two waves from the point B would be in similar phases when the distance $AC + CB$, or twice the thickness of the film, is equal to the wave length, or to any multiple of the wave length of light in the film. This turns out to be not the case. In the conditions described no light leaves the point B. To account for this Young assumed that the reflection of light takes place, in different circumstances, with or without change of sign. At one of the surfaces, say at the upper one, the light is passing from a denser to a rarer medium, and from the analogy of the reflection of sound, we may suppose it to be reflected there without change of sign. At the other surface it passes from a rarer to a denser medium, and we accordingly suppose it to be reflected there with change of sign. In this way the light which is reflected in the film has its phase reversed, and in considering its combination with the light reflected on the upper surface, we must consider its phase opposite to that which would be due to the distance over which it travels in the film. With this additional hypothesis, which subsequent analysis has shown to be justified, Young was able to explain the colors of thin films, and the size of the successive rings in Newton's experiment, and to show that the wave lengths of the different colors, calculated from the size of the rings, were in agreement with those obtained by him from his original experiments on interference.

If we trace the course of the waves which emerge after passing through the film, we find that, of the waves which are superposed at the point D, one of them has passed through the film once, while the other has passed through it three times, and has besides undergone two reflections with change of sign. The difference of phase of the waves superposed at D will therefore depend only on the difference of their paths in the film, and light will emerge when the difference in path is equal to an even number of half wave lengths in the film. Light is thus transmitted through the film when its thickness is such that no light is reflected by it. In Newton's experiment a system of rings will be seen by transmitted light which is complementary to those seen by reflected light; that is, in which the colors are so arranged that when they are superposed on the reflected system a uniform field of white light is the result.

A more extended study of the colors of thin films, in which account is taken of the fact that the light which enters the film does not all leave it after one or two internal reflections, confirms the

conclusions of the elementary theory in the case of the reflected light, but shows that the transmitted light will never be entirely extinguished for any thickness of the film. As the thickness of the film changes, the transmitted light will pass through a series of maximum and minimum values. These conclusions are confirmed by observation.

348. **Diffraction.** — Young also applied the wave theory to explain diffraction. He supposed that the diffraction bands outside the shadow are caused by the interference of the light which passes the obstacle outside its geometrical shadow, with light reflected from the edge of the obstacle. In this way he could explain in general the production of the diffraction bands, but he could not account for their exact position, nor could he readily explain some other of the diffraction phenomena.

349. **Fresnel's Demonstration of Interference.** — A few years after Young's discovery of interference, Fresnel (1788–1827) began researches by which the wave theory of light was highly developed and established. His early experiments on interference (1815) were similar to Young's, but he made an important step in advance by demonstrating the interference of light in a way that met an objection which had been raised to Young's experiment. In regard to that experiment it was said that the black bands observed by Young were not produced by simple interference, but were diffraction phenomena due to an action upon the light of the edges of the opening through which it passed. Fresnel avoided this objection by substituting for the two openings two images of the source formed in two plane mirrors, which met along one edge, and were very slightly inclined to each other. By properly adjusting the inclination of these mirrors, the two beams of light reflected from them were made to overlap, and interference bands, similar to those observed by Young, were produced. Fresnel found that by placing an eyepiece in the path of these beams the interference bands could be observed in it. He was thus able to observe them and to measure their distances much more precisely than can be done when they are allowed to fall on a screen. All the interference and diffraction phenomena can be observed best in this way.

350. **Rectilinear Transmission of Light.** — It was evident that the wave theory, however successful it might be in explaining some of the phenomena of light, could not be accepted if it could not be made to account for the rectilinear transmission of light. By a combination of Huygens' principle with Young's principle of inter-

ference, Fresnel was not only able to account for rectilinear transmission, but also to explain by the same principles all the phenomena of diffraction. To describe Fresnel's explanation of rectilinear transmission of a linear wave, we construct a diagram as follows (Fig. 199): Supposing the source of light to be a point at an infinite distance on the left, we draw a straight line to represent one of the rays sent from it, and at right angles to this line we draw a number of equidistant straight lines. The distance between two of these lines is taken to be half a wave length. The successive lines, therefore, represent parts of the successive waves which are in opposite phases. Take a point on the ray and with it as center describe a circle. By the principle of Huygens every point on this circle, being disturbed by the waves passing over it, will act as a center of disturbance and will send out a wave. The waves from all points on the circle will reach the center at the same instant, and their phases at that center will differ in the same way that they differ in the circle. The effect at the center will be that due to their superposition. Now if we examine those portions, or elements, of the circle which are intercepted between the parallel lines of the figure, we see that the disturbances from two successive elements will in general be in opposite phases, so that when they are superposed at the center, they tend to destroy each other. If we at first confine our attention to that portion of the circle which lies nearer the source, we see at once that much the largest element is that which lies nearest the line joining the source with the center of the circle, and that the successive elements, as we go out around the circle, soon become appreciably equal to one another. It is natural to suppose that the effect of any one of these elements is proportional to its length. Those elements which are nearly equal to one another, therefore, annul each other's effects in pairs, and the effect at the center is due to the very considerable preponderance of the effect of the first element over that of those which lie outside of it.

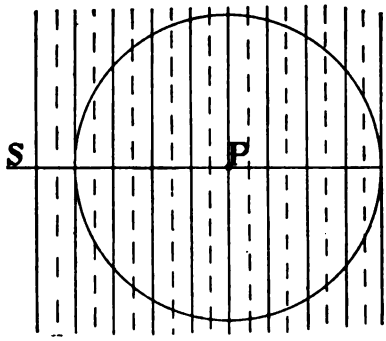


Fig. 199.

If we consider that portion of the circle which lies farther from the source than the center, we shall find the same difference in the

size of the successive elements as we found before. We may, however, neglect the action of this part of the circle if we remember that in the application of Huygens' principle we have agreed to suppose that the effect produced by the elementary waves diminishes with the obliquity, and that in particular their effect vanishes in the direction opposite to that in which the actual waves progress.

It thus appears that the effect which reaches the center of the circle is practically due entirely to that element of the circle which lies nearest the straight line joining the source with the center. The light received at the center, therefore, appears to travel from the source in a straight line. The conclusion which we have drawn from this description is fully confirmed by analysis.

351. **Diffraction at an Edge.** — Fresnel applied similar principles to the explanation of the diffraction produced by the edge of an obstacle. The linear waves which cause the effect are perpendicular to the edge as they pass the obstacle. To consider this case we construct the diagram as before, and suppose, first, that the obstacle is so interposed that the waves which would pass the circle are all intercepted. As the obstacle is lowered (Fig. 200) waves are set up at the different portions of the circle in succession. Those from the uppermost parts of the circle have no appreciable effect, on account of the obliquity, but as the obstacle is still further lowered, waves are at last set up in elements which send appreciable effects to the center of the circle. As we have already seen, the element nearest to the obstacle will be the largest, and it will send an effect to the center that will not be entirely annulled by the effects of the other elements. As the edge of the obstacle approaches the line from the source to the center, the element nearest to it becomes relatively larger and its effect relatively greater. When the edge of the obstacle is on that line, the effect at the center is that produced by half the wave. The wave theory, therefore, leads naturally to an explanation of the illumination of the geometrical shadow.

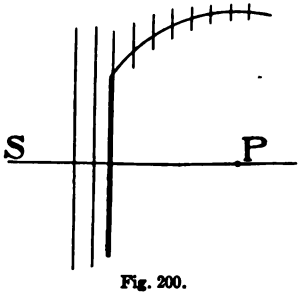


Fig. 200.

If the obstacle is still further lowered (Fig. 201) so as to expose the first element on the other side of the line joining the source with the center, the effect of that element is added to that of the upper half of the circle and the center is more highly illuminated. If the obstacle exposes two elements below the line, their effects

at the center, being in opposite phases, partly destroy each other and the center is relatively darker. Similar alternations of relative brightness and darkness occur at the center as the successive elements are exposed. The wave theory thus accounts for the diffraction bands seen outside the geometrical shadow.

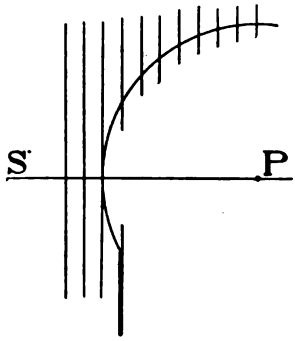


Fig. 201.

When Fresnel made a calculation of the positions of the diffraction bands, using the wave length of light which he had already obtained from interference experiments, he found that the calculated positions agreed with the observed positions of the bands.

352. Diffraction by a Slit. — We may consider also the diffraction produced by a narrow slit. If we suppose a slit set in front of our diagram, certain of the elements of the circle will be exposed to the light coming from the source, and we may determine the effect produced by the slit by considering the elements of the circle which will be exposed. If, for example, the slit is so placed as to expose the two central elements, the center of the circle will be brightly illuminated. If, on the other hand, it exposes the first and the second element, they will partly counteract each other and the center will be relatively darker. If it exposes three elements (Fig. 202), the third will supplement the effect of the first and the center will be brighter again. In general the center will be illuminated if the slit exposes an odd number of elements and will be dark if it exposes an even number.

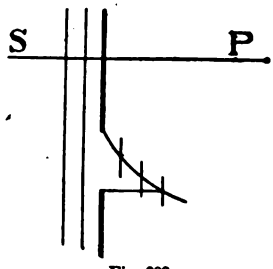


Fig. 202.

353. Colors of Diffraction Bands. — On the natural assumption that the color of light depends on its wave length, it is easy to explain the colors in the diffraction bands. For, because of the difference of wave length, the elements of our diagram will differ in length for the different colors, and a certain position of the slit which will produce darkness at the center for one wave length will not do so for another. For different positions of the slit, therefore, if we use white light, the center will be illuminated with different colors. These colors are generally due to the absence of one of the colors of the spectrum, owing to its forming a black band at the

center, and to the presence of the other colors in somewhat different proportions from those in which they occur in white light. By trials with lights of various colors it has been found that the wave length of the violet is the least, and that of the red the greatest, of the colors which constitute the spectrum.

354. The Diffraction Grating. — The principles which we have examined are exemplified by the instrument called the *diffraction grating*. It consists of a very large number of equidistant slits, made by cutting lines with a diamond upon a glass plate or upon a plate of speculum metal. It may be shown to be a consequence of the wave theory, and it has been verified by experiment, that when light of one wave length falls upon such a grating from a point source, the diffraction bands produced are very intense and narrow. The distance between them depends on the distance between the successive cuts on the grating, and may be made large by making the cuts near together. In the gratings used in most spectroscopic work, there are from 15,000 to 20,000 cuts to the inch, and the diffraction bands are so widely separated that only three or four of them can appear in front of the grating. When white light falls on such a grating, its constituent colors are diffracted at different angles, according to their wave lengths. With fine gratings, having a large number of lines, the spectrum thus obtained is of great purity. Since the distance between the diffraction bands produced by different wave lengths is proportional to the difference between the wave lengths, this spectrum is called a *normal spectrum*.

The diffraction grating furnishes the means of making accurate determinations of the wave lengths of light from various sources.

355. The Interferometer. — Another instrument, called the interferometer, which can be used for the determination of wave lengths, or for the measurement of distances in terms of wave lengths, was introduced in an elementary form by Jamin, and has been highly developed, and its usefulness much extended, by Michelson and others.

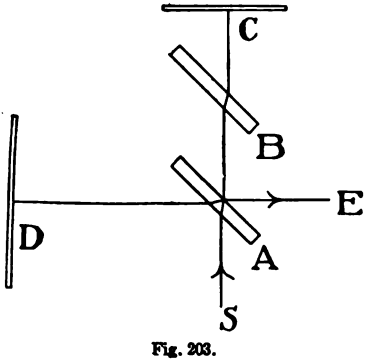


Fig. 203.

Michelson's *interferometer* is arranged in plan as follows (Fig. 203): Parallel light from a source S falls on a glass plate A with plane parallel sides, and slightly silvered on the back. Part of the light

passes through the silver coating and proceeds to the plane mirror C, passing through on its way the glass plate B, which is as exactly as possible similar to the plate A. The light falling on C is reflected from it back to A, and thence to the eyepiece at E.

Another part of the light which falls on the silvered surface of A is reflected directly to another mirror D, and is reflected from it and reaches the eyepiece E after traversing A again. The plate B is inserted to make the paths of the light optically similar, for the light which reaches the eyepiece by way of D traverses the plate A three times, while that which proceeds by way of C traverses it only once. By passing twice through the plate B, which is similar to A, this difference is corrected.

If the distances passed over by the two beams of light, after separation at the silvered surface of A, are accurately equal, the beams will blend in the eyepiece into one beam of maximum intensity. If the mirror D is moved out a half wave length by means of a micrometer screw, the beams will be so related as to interfere destructively, and the field of view is darkened. The field will become alternately bright and dark as the mirror D, in moving out, increases the path of the light which is returned to the eye from it, by half wave lengths.

The incident light is never strictly parallel, and the phenomenon usually seen is a series of interference fringes, often approximately arcs of circles, and resembling in general Newton's rings, which move across the field of view as the position of the mirror D is altered. When monochromatic light is used, the number of these which cross the center of the field of view for a certain displacement of the mirror D is a measure of the distance through which D has been moved in half wave lengths of the light employed.

If the wave length of a monochromatic light from some uniform source is taken as a standard, it is easy to see how a length may be measured by this instrument in terms of that standard. By an ingenious use of a series of intermediate standards, of which the first was directly measured in wave lengths by the interferometer, and the others compared successively by the same instrument, Michelson has determined the length of the standard meter in terms of the red, green, and blue rays from cadmium vapor. The numbers of wave lengths of these rays in one meter are respectively 1,553,163.5, 1,900,249.7, 2,083,372.1. In view of the probability that the wave lengths thus measured will be reproduced whenever light is set up at a similar source, this result furnishes a means

of controlling the length of the standard meter, or of reproducing it if, by any accident, it were to be defaced or lost.

The wave lengths of the particular waves thus measured are respectively, in millimeters, 0.000643847, 0.000508582, 0.000479991.

EXAMPLES, XIX

1. *To show that the mean kinetic energy of a particle vibrating with a simple harmonic motion is proportional to the square of the amplitude.*

a. The kinetic energy of the particle, of mass m, as it passes through the center of its path is $\frac{1}{2}mv^2 = \frac{2m\pi^2a^2}{T^2}$.

b. The potential energy at the end of the path is equal to the mean force multiplied by the amplitude. The acceleration at any point x is $4\pi^2x/T^2$ and the mean acceleration in the path from the end to the center is $2\pi^2a/T^2$. The potential energy required is therefore $2m\pi^2a^2/T^2$. This potential energy when the particle is at rest is equal to the kinetic energy at the center, as it should be.

c. The total energy, being constant, is always equal to this quantity. The mean kinetic energy is half this, and equal to the mean potential energy. This is seen most easily by expressing the velocity at any point by the help of the ordinate y of the auxiliary circle (§ 66). We have at any point the total energy equal to the sum of the kinetic and potential energies; or

$$\frac{2m\pi^2}{T^2}a^2 = \frac{2m\pi^2}{T^2}y^2 + \frac{2m\pi^2}{T^2}x^2,$$

since velocity $= \frac{2\pi a}{T}\sin\frac{2\pi t}{T}$, and $a\sin\frac{2\pi t}{T} = y$. The mean energies are therefore proportional to the mean values of x^2 and y^2, and these mean values are manifestly equal, from the relations of x and y in the circle. The mean value of the kinetic energy is therefore proportional to the square of the amplitude.

2. *On the assumption that the waves of light are transmitted in the ether without loss of energy, prove that the intensity of light is proportional to the energy which passes through a unit area perpendicular to the wave normal.*

Let e_1 represent the energy which passes through unit area in unit time at the distance r_1 from the source. Then the total energy which passes through the surface of the sphere of radius r_1 in unit time is $4\pi r_1^2e_1 = 4\pi r_2^2e_2$, the total energy which passes through the surface of a sphere of radius r_2. Now we know by experiment that the intensities i_1 and i_2 at the distances r_1 and r_2 from the source are connected by the relation $i_1 : i_2 = r_2^2 : r_1^2$, and so from the equation above, $i_1 : i_2 = e_1 : e_2$.

3. *If the mean kinetic energy of a vibrating ether particle is taken as the measure of the intensity of the light transmitted by it, show that the amplitude of light radiating from a center varies inversely as the distance from the source.*

The energy per unit area varies inversely with the square of the distance from the source (Example 2). It is also proportional to the square of the amplitude of the vibrations (Example 1). Hence the amplitude varies inversely as the distance from the source.

4. *To find the effect of two waves of the same period meeting at a point in different phases when the vibrations are in the same plane.*

Represent the displacements at the point by

$$y_1 = a_1 \cos 2\pi\left(\frac{t}{T} - \frac{x_1}{\lambda}\right), y_2 = a_2 \cos 2\pi\left(\frac{t}{T} - \frac{x_2}{\lambda}\right).$$

In this mode of representation the difference in phase is supposed to be due to a difference in path of the two waves. We have

$$y_1 = a_1 \cos 2\pi \frac{t}{T} \cos 2\pi \frac{x_1}{\lambda} + a_1 \sin 2\pi \frac{t}{T} \sin 2\pi \frac{x_1}{\lambda},$$

$$y_2 = a_2 \cos 2\pi \frac{t}{T} \cos 2\pi \frac{x_2}{\lambda} + a_2 \sin 2\pi \frac{t}{T} \sin 2\pi \frac{x_2}{\lambda}.$$

The resultant displacement is $y = y_1 + y_2$;

$$y = \cos 2\pi \frac{t}{T}\left(a_1 \cos 2\pi \frac{x_1}{\lambda} + a_2 \cos 2\pi \frac{x_2}{\lambda}\right) + \sin 2\pi \frac{t}{T}\left(a_1 \sin 2\pi \frac{x_1}{\lambda} + a_2 \sin 2\pi \frac{x_2}{\lambda}\right).$$

If we set

$$A \cos 2\pi \frac{\phi}{\lambda} = a_1 \cos 2\pi \frac{x_1}{\lambda} + a_2 \cos 2\pi \frac{x_2}{\lambda}, A \sin 2\pi \frac{\phi}{\lambda} = a_1 \sin 2\pi \frac{x_1}{\lambda} + a_2 \sin 2\pi \frac{x_2}{\lambda},$$

we may write $y = A \cos 2\pi\left(\frac{t}{T} - \frac{\phi}{\lambda}\right)$.

The motion is still simple harmonic. The amplitude A is obtained by squaring and adding the two equations which define A. We get

$$A^2 = a_1^2 + a_2^2 + 2a_1a_2 \cos 2\pi \frac{x_1 - x_2}{\lambda}.$$

The amplitude of the resultant vibration depends therefore on the difference of path of the two wave trains. If $x_1 - x_2 = (2n-1)\frac{\lambda}{2}$, we have $\cos(2n-1)\pi = -1$, $A^2 = (a_1 - a_2)^2$, and a minimum. If $a_1 = a_2$, as it is nearly in such experiments as Young's experiment, $A^2 = 0$, and the points on the screen for which the distances to the two sources differ by an odd number of half wave lengths are dark. When $x_1 - x_2 = 2n\frac{\lambda}{2}$, we have $\cos 2\pi n = 1$, $A^2 = (a_1 + a_2)^2$, and a maximum.

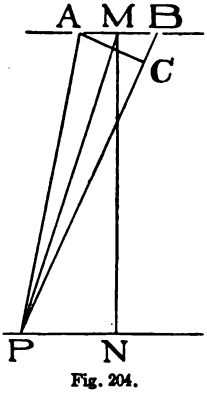


Fig. 204.

5. *To find the formula for the positions of the interference bands in Young's experiment.*

A black line will appear on the screen at P (Fig. 204) if the difference of length $BP - AP = (2n-1)\frac{\lambda}{2}$. When the point P is not far from the central point N of the pattern, this difference in length is approximately equal to BC, obtained by drawing AC perpendicular to BP, and AC is approximately equal to AB. We set $AB = s$, $PN = x$, $MN = a$, $BC = (2n-1)\frac{\lambda}{2}$. The triangle ABC is very approximately similar to the triangle MPN, so that we have

$$x : a = (2n-1)\frac{\lambda}{2} : s; \text{ or } x = (2n-1)\frac{\lambda}{2}\frac{a}{s}.$$

The distance of the successive black bands from one another is therefore a constant. The centers of the bright bands are located by

$$x = 2n\frac{\lambda}{2}\frac{a}{s}.$$

6. *To find the radii of Newton's rings in homogeneous light.*

The rings are formed between a plane surface and a convex surface of radius R (Fig. 205). The thickness d of the air film at a point distant r from the point of contact of the two surfaces is given by the geometry of the circle by $d = \frac{r^2}{2R - d}$. Since we consider only small distances from the center and since R is very large, we set $d = \frac{r^2}{2R}$. Now as described in § 347, interference occurs and no light is reflected when the difference in path of the two beams is equal to any multiple of a wave length, or when the thickness of the film is equal to any multiple of a half wave length. The black rings will appear, therefore, for such values of r that $d = \frac{r^2}{2R} = \frac{n\lambda}{2} = 2n\frac{\lambda}{4}$.

Fig. 205.

The bright rings will appear between the black ones, at such points that

$$d = \frac{r^2}{2R} = \frac{n\lambda}{2} - \frac{\lambda}{4} = (2n - 1)\frac{\lambda}{4}.$$

The radii of the successive black rings are therefore proportional to the square roots of the even numbers, those of the bright rings to the square roots of the odd numbers.

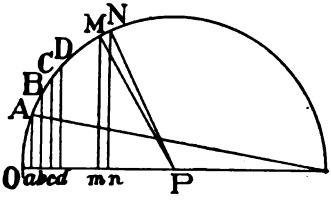


Fig. 206.

7. *To calculate the length of a half-period element of the circle represented in Fig.* 206.

These elements are so small that we may consider the arc of any one of them equal to its chord. Using R to represent the radius of the circle, we have, from a well-known proposition in geometry,

$$Oa : OA^2 = OA^2 : 2R,$$
$$Ob : OB^2 = OB^2 : 2R, \text{ etc.},$$

and since $Oa = \frac{\lambda}{2}$, $Ob = 2\frac{\lambda}{2}$, etc., we have

$$OA = \sqrt{R\lambda},\ OB = \sqrt{2R\lambda},\ OC = \sqrt{3R\lambda},\ OD = \sqrt{4R\lambda}, \text{ etc.}$$

Hence $OA = \sqrt{R\lambda}$, $AB = \sqrt{R\lambda}(\sqrt{2} - \sqrt{1})$, $BC = \sqrt{R\lambda}(\sqrt{3} - \sqrt{2})$, $CD = \sqrt{R\lambda}(\sqrt{4} - \sqrt{3})$, etc., or $OA = \sqrt{R\lambda}$, $AB = 0.414\sqrt{R\lambda}$, $BC = 0.319\sqrt{R\lambda}$ $CD = 0.266\sqrt{R\lambda}$, etc. The differences of the successive elements decrease very rapidly.

When the element is one like MN, which is at some distance on the circle from the point O, it may be very approximately represented by $MN = \frac{\lambda}{2\sin\phi}$, using ϕ to represent the angle MPO. As the sine of a fairly large angle increases only very slowly as the angle increases, the successive elements in this part of the circle will decrease very slowly, and their differences will be very small.

8. *To find the effect of a train of plane waves at a point.*

If we construct the same diagram as that used in Example 7, and consider it rotated about the line SP as an axis, the elements of the circle will describe circular zones. The disturbances which leave these zones at the same instant will meet at the center P, and the effect at P will be due to their superposition.

The area of any zone, such as the one of which BC is the chord, is $2\pi R \cdot bc$. Now $bc = \frac{\lambda}{2}$; and the same is true for the heights of all the zones. Hence each zone has the same area, and if the effects of the zones at P depend only on their areas, we should conclude that the disturbances from the successive zones, being in general in opposite phases, would annul each other, and that the center P would be left undisturbed. This conclusion is avoided by assuming that the effect of each zone, while proportional to its area, depends also upon the direction in which the elementary waves leave the wave front. It is assumed that the effect diminishes as the angle made with the wave normal by the direction in which the disturbance proceeds increases. Concisely stated, it may be said that the effect diminishes with the obliquity. We may then, following Fresnel, represent the effect at the center P by the series

$$m_1 - m_2 + m_3 - m_4 + \cdots \pm m,$$

in which each term represents the effect of a zone. The first term is the greatest and the others follow in the order of their magnitudes. This series may be written

$$\tfrac{1}{2}m_1 + (\tfrac{1}{2}m_1 - \tfrac{1}{2}m_2) - (\tfrac{1}{2}m_2 - \tfrac{1}{2}m_3) + (\tfrac{1}{2}m_3 - \tfrac{1}{2}m_4) -, \text{etc.}$$

The sum of the successive terms in parenthesis vanishes in comparison with the first term. This seems probable on inspection and can be justified by a study of the series. The total effect at the center P is therefore equal to half that of the first element. The light appears to reach the point P by traveling along the line joining it with the source S.

9. *To calculate the effect of a linear wave at a point.*

For most purposes it is best to consider the effect at a point as due to the superposition of disturbances which leave a straight line in space parallel with the wave fronts at such different times that they all arrive at the illuminated point together (Fig. 207). To obtain the effect, increase the distance $PO = a$ by successive increments of half a wave length, and with these lines as radii describe circles, about P as a center, which cut the chosen line OY at the points A, B, C, . . . A disturbance which leaves A will reach the point P together with one which leaves O at a time later by half a period, and these two disturbances will therefore be in opposite phases. In the same way disturbances which leave the points B, C, etc., and reach the point P together, will be in opposite phases. In general, the disturbances which reach P from the element OA will be in opposite phase to those which reach P from the element AB; and the effects from the successive elements will be alternately, as we may say, positive and negative in phase. We obtain the whole effect at P by assuming that the effect of each element will be proportional to its length, and

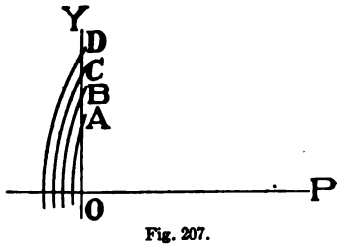


Fig. 207.

by adding the lengths of the elements taken alternately with positive and negative signs.

From the mode of construction, setting $OA = y_1$, $OB = y_2$, $OC = y_3$, etc., we have

$$\left(2a + \frac{\lambda}{2}\right)\frac{\lambda}{2} = y_1^2, \quad \left(2a + \frac{2\lambda}{2}\right)\frac{2\lambda}{2} = y_2^2, \quad \left(2a + \frac{3\lambda}{2}\right)\frac{3\lambda}{2} = y_3^2, \text{ etc.},$$

or, since λ is very small compared with a, so that we may omit the term in the parenthesis containing λ,

$$y_1 = \sqrt{a\lambda}, \quad y_2 = \sqrt{2a\lambda}, \quad y_3 = \sqrt{3a\lambda}, \quad \text{etc.},$$

and the lengths of the elements are

$$y_1 = \sqrt{a\lambda}, \quad y_2 - y_1 = \sqrt{a\lambda}\,(\sqrt{2} - \sqrt{1}), \quad y_3 - y_2 = \sqrt{a\lambda}\,(\sqrt{3} - \sqrt{2}), \text{ etc.}$$

The lengths of the successive elements decrease as we go out from O, more and more slowly as we go farther out. The effect is, as shown in the similar case in Example 7, approximately equal to that of half the first or central element.

This method is especially convenient when we have to deal with the waves passing through small openings.

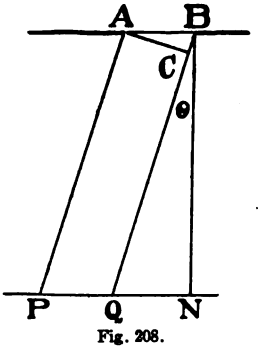


Fig. 208.

10. *To calculate the position of the diffraction bands produced by a plane wave falling on a narrow slit.*

We can treat this by considering a linear wave transverse to the slit. The points in the slit AB (Fig. 208) lie on certain half-period elements of the wave. If the distance BC is an even number of half wave lengths, the slit will admit an even number of elements, and they will counteract each other in the direction AP. There will therefore be a dark band on a screen where P or Q meets it, the two points being so near together that they may be treated as coincident. To locate Q we have $BC = AB \sin\theta$, and $QN = QB \sin\theta$; so that

$$QN = BC\frac{QB}{AB}.$$

To generalize this formula we set $QN = x$, $BC = 2n\lambda/2$, $AB = s$, and $QB = BN = a$ very approximately, since x is always small, and have

$$x = 2n\frac{\lambda}{2}\frac{a}{s}.$$

The light bands will occur at points on the screen for which BC equals an odd number of half wave lengths, for which

$$x = (2n - 1)\frac{\lambda}{2}\frac{a}{s}.$$

11. *To calculate the position of the bright bands formed by a diffraction grating.*

We assume that the opaque lines of the grating are of the same width as the transparent spaces between them. A certain direction AP (Fig. 209) will exist for which the distance Ab is equal to a wave length. The distance AB will then contain two elements, one of which is prevented from affecting the region beyond the grating by coinciding with the opaque line. The element which is not obstructed will send light in the direction AP. For the same reason each of the

openings of the grating will send light in the same direction. The light coming from these openings proceeds from every other half-period element and so is in general in the same phase; that is, it is all positive in phase or negative in phase at the same time. The effect is then that of a beam of parallel rays proceeding in the direction AP.

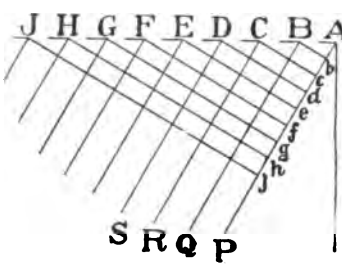


Fig. 209.

A similar beam will proceed in another direction, more inclined to the face of the grating, for which the distance Ab is equal to two wave lengths; and so for other inclinations, for which Ab equals any other number of wave lengths.

The inclination $PAN = \theta$ is obtained, if we set $AB = s$, usually called the width of an element of the grating, from the formula $n\lambda = s \sin \theta$.

Since gratings with a large number of elements give very sharply defined positions of the lines corresponding to different values of λ, and since s and θ can be accurately measured, the value of λ can be determined most satisfactorily by means of the diffraction grating.

Polarization

356. Luminous Vibrations Transverse. — In the form of the wave theory first used by Fresnel, the waves were thought of as longitudinal. That is, the vibrations of the medium were supposed to take place to and fro in the line of transmission. An example of such a mode of vibration was already known in the case of sound, and it was also known that a fluid, as the ether was then assumed to be, could only vibrate in that manner. When Young and Fresnel tried to apply the wave theory to the explanation of double refraction and polarized light, they found it impossible to make any headway so long as they retained the hypothesis of longitudinal vibrations. In fact it is clear that such vibrations cannot present any such distinct differences on different sides of the ray as occur in the case of polarized light. Both Young and Fresnel determined, therefore, to abandon the hypothesis of longitudinal vibrations, and to adopt in its stead the hypothesis that the vibrations are more or less transverse to the line of progress. Young did little more than indicate his adoption of this view, and its development was entirely due to Fresnel.

In order to verify the hypothesis of transverse vibrations, so far as it can be done by experiment, Fresnel and his friend Arago (1819) executed a series of experiments on the interference of polarized

light. To understand the statement of their results, it must be mentioned that the two polarized beams which pass through a crystal of Iceland spar are said to be polarized in opposite planes, meaning thereby in planes at right angles to each other but parallel with the direction of the beam. Fresnel and Arago tried Young's experiment to obtain interference, only using polarized light, variously modified, instead of natural light. They found that when a beam of polarized light fell upon the source and was not modified before it fell upon the two openings, the interference phenomena obtained were just the same as those obtained with ordinary light. By interposing properly prepared crystals in the paths of the polarized light falling from the source on the two openings, they polarized the beams which fell on the openings in opposite planes, and then found that interference did not occur. This result is in accord with the hypothesis that the vibrations of light are transverse to the line of progress, and are at right angles to each other in oppositely polarized beams. For, when the beams from the two openings in the first experiment were similarly polarized, they would be in the same plane and so could interfere destructively. When, in the second experiment, the two beams were polarized in opposite planes, the vibrations would be perpendicular to each other, and so could never destroy each other by interference; for it is plain that two vibrations at right angles to each other can never act on one particle so as to keep it at rest. Fresnel concluded from these experiments that the hypothesis of transverse vibrations was confirmed.

357. **Common Light.** — It may be well to consider at this point Fresnel's description of common light on the hypothesis of transverse vibrations. He supposed monochromatic light to be such a motion in the medium as may be obtained by the superposition of two simple harmonic motions transverse to the line of progress and at right angles to each other. The path of a point describing such a motion is, in general, an ellipse. Two such motions may be superposed so as to produce interference so long as the elliptic paths are similar. The fact that interference can be obtained between two rays of light which differ in length by 2,500,000 wave lengths, shows that the vibrations at the source remain similar, and send out similar disturbances through space, for a time containing at least 2,500,000 periods of the vibration. On the other hand, it is also evident that the vibration from a source does not remain always the same. If it did do so, it would be polarized, and the two rays

into which it is broken by a doubly refracting crystal would differ in intensity from each other, and would have different relative intensities for different positions of the crystal. Now observation shows that both the rays transmitted by the crystal have the same intensity when the light used is common light. We can explain this only by supposing that the phase of one of the component vibrations, or the phases of both of them, change abruptly from time to time, so as to alter the polarization of the vibration. Since over 500 million million vibrations are executed by yellow light in one second, a sufficient number of such changes may occur in that time to give the two component vibrations into which the common light is divided by the crystal equal average intensities, and yet those vibrations may continue in one phase long enough to account for the interference of rays whose paths differ by millions of wave lengths.

This description of common light is confirmed by another experiment of Fresnel on the interference of polarized light. He found that if light originally polarized before it entered the source was divided into two beams polarized in opposite planes, which fell on the two openings, these beams, which, in this condition, would not interfere, could be made to interfere if their planes of polarization were brought into coincidence. On the contrary, if the light which fell on the source was not polarized, and if its two beams which fell on the openings were first polarized in opposite planes, the beams thus formed did not interfere, even when they were brought into the same plane of polarization. In the latter experiment we notice that the two polarized beams which fall on the two openings contain the two components of the elliptic vibration of common light. According to our hypothesis, these components change their phases occasionally, and there is no reason to think that their changes of phase will always occur together. If, therefore, those components of them which fall in the similarly polarized beams upon the receiving screen are at one instant so related as to interfere at one point, they will not remain so long enough for the interference to be perceptible. The interference bands will pass from one point to another on the screen so many times in a second that the illumination of the screen will appear uniform.

358. Polarization by Reflection. — Fresnel employed the hypothesis of transverse vibrations to explain *polarization by reflection.* It had been discovered a few years before by Malus (1808) that when light is incident upon a reflecting surface, the light in both the reflected and the refracted beams is generally partially polarized. For a certain angle of incidence, called the *polarizing angle*, the reflected light is completely polarized. Most reflecting substances, except the metals, have this property of polarizing light by reflection. It was discovered by Brewster (1815) that the most complete

polarization occurs when the angle of incidence (Fig. 210) is such that the reflected ray and the refracted ray are at right angles to each other, or when the tangent of the polarizing angle is equal to the index of refraction. The planes of polarization in the reflected and the refracted beams are perpendicular to each other. When the polarization is complete in the reflected beam, the polarization is a maximum in the refracted beam.

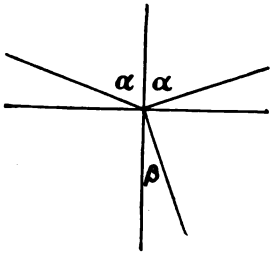


Fig. 210.

By making certain suppositions regarding the relations of the components of the vibrations at the reflecting surface, which were not entirely in accord with mechanical principles, although he treated the vibrations as if they were the vibrations of some sort of matter, and as if they conformed to the law of conservation of mechanical energy, Fresnel was able to show that if the angle of incidence is the polarizing angle, light transmitted by vibrations occurring in the plane of incidence will not be reflected, but will be entirely refracted. Light transmitted by vibrations at right angles

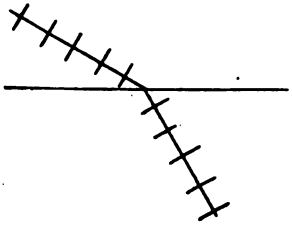


Fig. 211a.

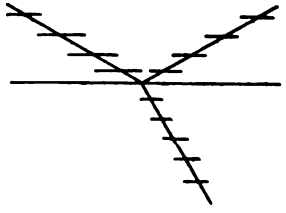


Fig. 211b.

to the plane of incidence will be partly reflected and partly refracted (Figs. 211a, 211b). The effect of reflection at the polarizing angle upon common light is therefore to reflect a portion of the light corresponding to the component of its elliptic vibration which is perpendicular to the plane of incidence, while the light corresponding to the other component, lying in the plane of incidence, is contained only in the refracted beam (Fig. 212). At any other angle of incidence than the polarizing angle, both these components occur in both the reflected and the refracted beams, but in

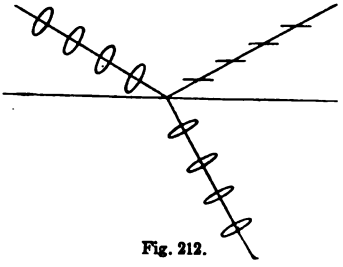


Fig. 212.

different proportions, so that those beams are partially polarized. Fresnel's theory led to certain relations between the intensities of the reflected and the refracted beams, which he found to be in agreement with observation.

359. Polarization by Double Refraction. — When polarized light was first studied in connection with double refraction, the plane of polarization was specified by reference to a definite plane in the crystal. A typical crystal of Iceland spar is a rhombohedron (Fig. 213) bounded by six equal faces, each of which is a similar rhombus. The line drawn from one obtuse angle A of this rhombohedron to the opposite obtuse angle B is called the *axis* of the crystal. It is the line marking the direction along which no double refraction takes place. We have already called it the *optic axis*. Any plane CD perpendicular to one of the faces of the crystal and parallel to the optic axis is called a *principal plane*. The ordinary beam emerging from the crystal is conventionally said to be polarized in the principal plane. The extraordinary beam, which is polarized oppositely to the ordinary beam, is then said to be polarized in a plane at right angles to the principal plane. The *plane of polarization* of any polarized beam is therefore the plane which corresponds to the properties of the beam in the same way that the principal plane of the crystal corresponds to the properties of the ordinary beam.

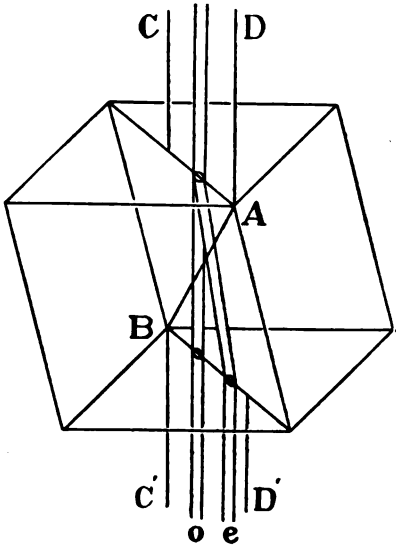


Fig. 213.

When a polarized beam falls upon a reflecting surface at the polarizing angle and in such a way that its plane of polarization is perpendicular to the plane of incidence, it is not reflected. According to Fresnel's hypothesis, therefore, its vibrations are in the plane of incidence. Fresnel concluded that the vibrations of the polarized beam are perpendicular to the plane of polarization.

By modifying one of Fresnel's hypotheses, F. Neumann and McCullagh developed a theory of polarized light which differed from

that of Fresnel in supposing that the vibrations are in the plane of polarization. Both theories were able to account for all known facts of observation. For many years they stood as alternative theories. They have been explained and reconciled by the electromagnetic theory of light.

Fresnel was able to explain the double refraction in Iceland spar and to develop a general theory of double refraction, which applies to all sorts of crystals, on the hypothesis of transverse vibrations. To do this he studied the effect of a disturbance set up inside a body, like a crystal, in which the elasticity is different in different directions. He showed that any general elliptic vibration will be resolved into two vibrations at right angles to each other, and that these will be transmitted in different directions in the crystal with different velocities. The surface reached at any instant by the vibrations which pass out in all directions from the disturbed center is what is called the *wave surface*. In most crystals, the wave surface is a complicated one, formed of two sheets, which touch each other at four symmetrically arranged points. These points may be connected in pairs by two lines which pass through the center. Four tangent planes may be drawn, parallel with one another in pairs, which touch the wave surface in circles. The normals to these planes are called the *optic axes* of the crystal, and represent directions in which plane waves may advance without exhibiting polarization or double refraction. For certain classes of crystals, of which Iceland spar is an example, the wave surface reduces to a sphere and an ellipsoid of revolution (Fig. 214), which touch each other at the ends of one of the axes of the ellipsoid. The line joining these points of contact is the optic axis. Crystals of this sort are called *uniaxial* crystals; crystals of the other sort are *biaxial*.

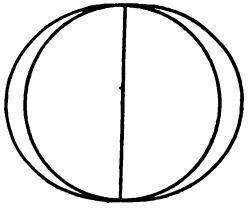


Fig. 214.

By the use of Fresnel's wave surface, the directions of the two rays formed by double refraction may be calculated for different angles of incidence, and the theory may thus be tested. The most minute observation has found no point of disagreement between the conclusions of the theory and the results of observation. The form of the wave surface in crystals is thus proved to be that developed by Fresnel. This result does not, however, confirm the hypotheses from which Fresnel deduced it. Other forms of the theory lead to the same, or practically to the same, form of the wave surface.

360. Chromatic Polarization. — The hypothesis of transverse vibrations received additional confirmation from its success in

explaining certain phenomena discovered by Arago. Arago found that if a parallel beam of polarized light was received upon a crystal so placed as to extinguish the beam, the light could be made to reappear by interposing in the path of the beam a thin sheet of mica. The light which thus appeared was colored, and the color changed as the crystal through which it was observed was rotated.

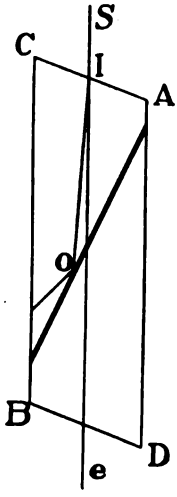


Fig. 215.

To show this experiment we use two crystals of Iceland spar, called Nicol's prisms. A *Nicol's prism* (Fig. 215) is a long prism of Iceland spar which has been cut through diagonally from one obtuse angle to the other, and cemented together again with Canada balsam after the new faces have been polished. A beam of light which falls on one end of this crystal is divided into two oppositely polarized beams, which proceed through the crystal till they meet the surface of the Canada balsam. If the crystal has been divided in the proper way, the ordinary ray will meet the balsam at an angle greater than its critical angle. It will consequently be totally reflected and will not emerge from the other end of the crystal. The extraordinary ray, on the contrary, meets the balsam within its critical angle, and is therefore partly transmitted through it, and emerges from the crystal. The Nicol's prism may thus be used to obtain a beam of polarized light.

In trying Arago's experiment, one Nicol's prism is used as the *polarizer;* the other, called the *analyzer*, is placed in the path of the polarized beam and the light which comes through it is observed either on a screen or by the eye. When the analyzer is turned so that its principal plane is at right angles to that of the polarizer, no light passes through it. If we now interpose between the two prisms a sheet of mica, colored light passes through the analyzer. In only two positions of the mica sheet will this not occur, when its principal plane is either parallel with, or perpendicular to, the plane of polarization of the incident beam. For other positions of the mica, no position of the analyzer can be found in which light will not pass through it.

Fresnel explained this phenomenon in the following way: Mica being a doubly refracting crystal, the vibrations of the polarized beam which fall on the mica sheet are resolved into two components perpendicular to each other (Fig. 216). The velocities of these components are different, and when they emerge from the mica one of them has gained a fraction of a wave length on the

other, so that they are then in different phases. Because the wave lengths of different colors are different, the differences in phase between the two emergent beams are different for the different colors. After emergence, these beams proceed to the analyzer, which resolves each of them into two perpendicular components, and transmits one component of each. The components transmitted are parallel with each other, and may therefore interfere. If they happen to be in opposite phases, they destroy each other. If they are in the same phase they enhance each other. For a given thickness of the mica some of the constituents of the original white beam will thus be destroyed, and the light which will pass will appear colored. The light which is rejected by the analyzer contains those colors which are absent in the transmitted light, so that if the analyzer is turned so that it transmits the light which it formerly rejected, the complementary color appears.

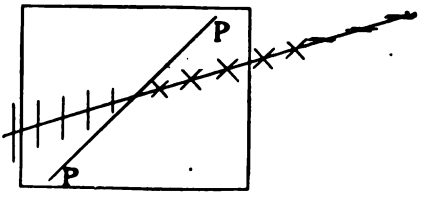


Fig. 216.

Similar effects may be produced by the use of thin sheets of crystals other than mica.

By using convergent light instead of parallel light, very complicated systems of colored figures may be produced. The peculiarities of these systems may be calculated from a knowledge of the optical properties of the particular crystal which is interposed in the beam. When the results of calculation are compared with the results of observation, a very complete agreement is found between them. This general result furnishes additional confirmation of Fresnel's wave surface.

361. Rotation of the Plane of Polarization. — Arago discovered a peculiar effect produced by quartz on polarized light, which was also explained by Fresnel. Quartz is a uniaxial crystal, and light falling perpendicularly upon a plate of quartz cut perpendicularly to the optic axis is not doubly refracted. If, however, an analyzer is so placed in a polarized beam as to extinguish it, and if then such a plate of quartz is interposed in the polarized beam, light will again come through the analyzer. Unless the plate of quartz is too thick, this light is colored. It is not altered by rotation of the quartz around the beam as an axis, but when the analyzer is rotated, the color changes continually.

If light of one color is used in this experiment, it will pass through the analyzer when the quartz plate is introduced, and if the analyzer

is then turned through a certain angle, which is different for each color, the light will be extinguished. From this experiment it is easy to see that the different colors which appear, when white light is used, are due to the suppression of some of the constituents of white light in each position of the analyzer. This action of quartz is known as the *rotation of the plane of polarization.* Many other substances, among them solutions of sugar and of other organic bodies, were subsequently found to possess the same property. From the fact that it is possessed by solutions, in which it is impossible to suppose that the active molecules have any definite directions, such as they may be supposed to have in crystals, we conclude that this action on polarized light is due to the structure of the molecule itself, or perhaps of one of its atoms.

Fresnel explained this phenomenon by supposing that the polarized beam, on entering the quartz, sets up two circular vibrations in opposite senses (Fig. 217). It is easy to see that the resultant of two such vibrations will be a rectilinear vibration like that of the incident beam. He further supposed that the velocity in the crystal of one of these vibrations is greater than that of the other. If this is the case, the circular vibration which is traveling more slowly, on its emergence from the quartz, will combine with the circular vibration which is traveling more rapidly, and which therefore enters the quartz at a later time, at which the moving element is further around on the arc of the circle which it is describing, to produce a rectilinear vibration, which is inclined to the one which entered the quartz by an amount proportional to the thickness of the plate. If the circular vibration which is traveling faster is the one in which the rotation is clockwise to an observer looking along the beam, the plane of polarization of the emergent light will be turned clockwise. If the other circular vibration travels faster through the quartz, the plane of polarization will be turned counterclockwise. Specimens of quartz are found which show each of these rotations. They are called right-handed and left-handed, respectively. By an ingenious combination of quartz prisms properly cut, Fresnel was able to separate the two circularly polarized beams assumed by him in this explanation, and so to prove its correctness.

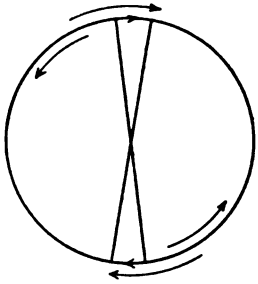


Fig. 217.

362. Elliptically and Circularly Polarized Light. — In his study of the reflection of polarized light Fresnel was led to consider the effect of total reflection upon the plane of polarization. In that case he found that when the vibration in the polarized beam is not at right angles to the plane of incidence, that component of it which lies in the plane of incidence has its phase reversed by the reflection. In general the combination of the new component thus produced with the other component which is reflected without change produces an elliptic vibration. By combining two such total reflections, taking place at the proper angles, the vibration of the reflected beam becomes circular. The beam is then said to be *circularly polarized.* Circular polarization may also be produced by the use of a sheet of mica, whose thickness is such that one of the two rays formed in it by double refraction gains a quarter of a wave length on the other in its passage through the sheet. This sheet is placed in the path of a polarized beam in such a position that its principal plane makes an angle of 45° with the plane of polarization. In this case the two components which emerge from it contain vibrations of equal magnitude and differing in phase by one quarter of a period. These two vibrations combine to produce a circular vibration. Such a sheet of mica is called a *quarter-wave plate.*

When polarized light is incident obliquely upon a polished metallic surface, the reflected light is elliptically polarized.

363. Magnetic Rotation of the Plane of Polarization. — It was discovered by Faraday (1845) that when a beam of plane polarized light traverses a magnetic field in the direction of its lines of force, the plane of polarization is rotated. The amount of rotation depends upon the strength of the field and upon the substance through which the light is passing. It was found by Faraday first in a glass of a peculiar composition. The rotation produced by the magnetic field differs from that produced by quartz in one important respect. If the beam which has been rotated by the magnetic field is received on a plane mirror and sent back through the field again, it undergoes an additional rotation in the same sense. On the other hand, a beam which is sent back through the quartz undergoes rotation in the opposite sense, so that its plane of polarization becomes the same as that of the incident beam. Magnetic rotation of the plane of polarization is explained in the same general way as the rotation by quartz. The two circularly polarized beams have been separated by Brace.

The Spectrum

364. Spectrum Analysis. — When salts of the elements are introduced into a colorless flame, like that of the Bunsen burner, they sometimes give the flame characteristic colors. For example, sodium chloride or sodium carbonate will color the flame yellow. If the light from this flame passes through a narrow slit and is received on a lens and prism so adjusted as to give a pure spectrum, certain parts of the spectrum are found to be much more intense than the rest of it. In the case of the sodium salts a narrow band or line, very brilliantly illuminated, appears in the yellow. This yellow line of sodium was noticed by Herschel. By the investigations of Bunsen and Kirchhoff (1859–1862) it was shown to be characteristic of the presence of sodium vapor in the flame, so that whenever this line can be detected, it may be inferred that sodium is present. With sufficiently high dispersion, this line is resolved into two lines, standing near together, and of about equal brightness. Bunsen studied the characteristic lines, or, as we may say, the spectra of different elements, and showed that they can be used as a means of detecting the presence of those elements in the substance by which the flame is colored. Elements which cannot be vaporized in a flame may be vaporized by the heat of the electric arc, and will then give similar characteristic spectra. When the electric spark is passed through an elementary gas, a similar characteristic spectrum is produced.

This method of detecting the presence of a particular element in a compound by means of its characteristic spectrum is called *spectrum analysis.*

365. Line, Band, and Continuous Spectra. — The spectrum of a gas or luminous vapor generally shows a set of characteristic lines, such as those which have been described. It is then called a *line* spectrum. In many cases it shows also sets of bands of light, which appear continuous with small dispersion, but which higher dispersion often resolves into very many fine lines. Such a spectrum is called a *band* or a *channeled* spectrum.

In some cases in which the spectrum of a gas has been observed as the pressure upon the gas is increased, it has been found that the lines, which at first are sharp and narrow, gradually broaden out into bands, and as the pressure is still further increased, these bands overlap, until the spectrum becomes continuous.

The spectrum of an incandescent liquid or solid is generally

continuous, showing no abrupt differences of intensity, and exhibiting every gradation of color from the extreme red to the extreme violet.

366. **Fraunhofer's Lines.** — When a pure spectrum is formed with sunlight, it is found to be crossed by a great number of dark lines. Some of the most intense of these lines were observed by Wollaston (1802), and they were more accurately studied by Fraunhofer (1814). The light from an incandescent body does not show these lines.

It was noticed by several observers that one of the most conspicuous of the *Fraunhofer lines*, in the yellow light of the solar spectrum, coincided in position with the yellow line of the spectrum of sodium. This coincidence was verified by the exact observations of Kirchhoff, who was working with Bunsen on the development of spectrum analysis. Kirchhoff suspected that the dark line in the solar spectrum is due to the absorption of light, coming from the central part of the sun, by the vapor of sodium in its outer atmosphere. To show that this may be the case, he observed the spectrum of the white light coming from incandescent lime, heated by the oxyhydrogen flame, when a flame containing sodium vapor was placed in front of the slit. He found, with those conditions, that a dark line appeared in the spectrum, coinciding in position with the bright yellow line which the sodium vapor itself would have given. He explained the production of this dark line by the aid of the principle of resonance. From the fact that sodium vapor emits light of a certain period, as shown by its giving rise to the characteristic yellow line of its spectrum, it is evident that those elements of sodium vapor which emit light execute vibrations of that period. Now, if light of all periods of vibration falls upon the sodium vapor, those vibrations which are not similar to the natural vibrations of the sodium will pass on without disturbing it, but those vibrations which have the same period as that of the sodium will increase its natural vibration by giving to it impulses properly timed, and so will themselves be diminished in intensity, or absorbed. The dark line which is thus cast in the spectrum is not black, but simply not so intensely illuminated as the regions on either side of it.

A comparison of the Fraunhofer lines of the solar spectrum with the spectra of the various elements led to the discovery of many coincidences similar to that described in the case of sodium. In some cases, as in that of iron, for example, these coincidences extend

to hundreds of lines. In every case in which such coincidences can be proved, we infer the presence of the particular element giving the spectrum in the outer atmosphere of the sun.

In Fig. 218 are represented some of the most conspicuous of the Fraunhofer lines. These are to be thought of as black lines ap-

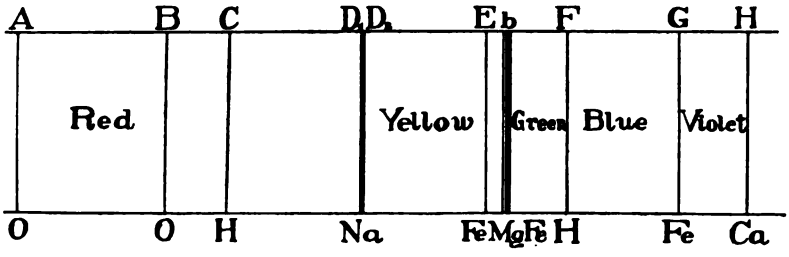


Fig. 218.

pearing on a colored ground. Below the diagram are indicated the elements in the spectra of which bright lines appear which coincide in position with the dark lines of the solar spectrum.

367. Absorption Spectra. — When a colored transparent body is placed in the path of a beam which forms a spectrum, there often appear particular portions of the spectrum in which the light is less intense than it was before. These darker regions are called *absorption bands*. They are characteristic of the particular substance which intercepts the light, and may be used as a means of analysis.

368. Kirchhoff's Law. — Kirchhoff found by a general theoretical investigation that the absorption of yellow light by sodium vapor is an example of a perfectly general law, that any body will absorb those waves of light which it will itself emit when self-luminous. When the emission of the radiation is due simply to elevation of temperature, the ratio between the emissive power and the absorptive power is the same for all substances at the same temperature. Those bodies which absorb all colors also emit all colors when they become self-luminous. Those bodies which absorb only particular colors emit only those particular colors.

369. Color of Bodies. — We are now in a position to consider the question of the *color of bodies*. All bodies which are not self-luminous are seen by means of the light reflected by them, coming from the sun, or from some other self-luminous source. Very many bodies show the same color when examined by transmitted light as when examined by reflected light. To explain their color, we suppose that some of the constituents of the white light which falls

upon them are absorbed, and that the light which is reflected to the eye, and by which the body is seen, has penetrated sufficiently within the body to allow this absorption to deprive it of those constituents. We therefore see it really by light which has traversed enough of the body for absorption to have its full effect.

There are many other bodies, however, for which the color in the transmitted light is different from that in the reflected light. With them the reflection seems to occur at the surface. It is found in every such case that absorption bands appear in the transmitted light, and that the light which is reflected is exactly that which is wanting in the transmitted beam. These bodies show also another peculiarity, which was discovered by Le Roux (1862) in iodine vapor and by Christiansen (1870) in fuchsine, and which was fully studied by Kundt. This peculiarity is called *anomalous dispersion.* In very many cases the spectrum formed by a prism of a particular substance, such as glass, for example, has the colors arranged in the order of their wave lengths. A dispersion of this sort is supposed to be according to law, and any deviation from it is anomalous. When light is sent through a prism of the substance showing anomalous dispersion, the order of the colors of the spectrum is not the order of the wave lengths. It was found by Kundt to be a general law that the colors which are displaced from their position in the ordinary spectrum are those which lie on either side of an absorption band. The color whose wave length is longer is displaced toward the violet end of the spectrum, and that whose wave length is shorter, toward the red end.

All these peculiarities of substances which show surface color and anomalous dispersion can be explained by supposing that the elements of their structure which can emit light have vibrations of their own of the same period as the light which the substance absorbs.

370. Stefan's Law. — The radiation from the bulb of a thermometer was investigated by Dulong and Petit, and was found to increase with the temperature. Stefan, by a more careful examination of the results, was able to state them in the law that the radiation was proportional to the difference of the fourth powers of the absolute temperatures of the thermometer and of the enclosing wall. Since by Prevost's law of exchanges (§ 254) each of these bodies should be supposed to be radiating to the other according to the same law, Stefan's law may also be stated by saying that the radiation from a body is proportional to the fourth power of its absolute temperature.

The many departures from this law which have been found by more extended observations are generally believed to depend upon the character of the radiating surface. The law is believed to hold accurately for a *black body*, that is, for a body which completely absorbs all the radiation of every wave length which falls upon it. In fact, a more satisfactory definition of a black body would perhaps be that it is a body which conforms at all temperatures to Stefan's law.

371. **The Extent of the Spectrum.** — Friedrich Wilhelm Herschel (1800) examined the heating effect produced by the different parts of the spectrum by placing a thermometer in it, and found that as the bulb of the thermometer was moved down toward the red end of the spectrum the heating effect became more and more pronounced, and that a still greater heating effect appeared when the bulb was placed in the region just beyond the red end of the spectrum. He concluded that rays exist of longer wave length than the red rays. When the action of light on salts of silver was discovered, on which photographic processes depend, it was found that the most active region in this respect lay outside the violet end of the spectrum. The invisible rays which produced the heating effect were at first called heat rays, and those which produced the chemical effect, actinic rays; but it is easily seen that there is no reason for considering them to be essentially different from the visible rays. Accordingly it has been found that all these rays, without exception, produce a heating effect, and the chemical effect has been caused by so many of them that the conclusion is fully warranted that the rays from a luminous body, or from any body, are of the same nature, and differ only in the lengths of the waves whose direction of transmission they indicate. The examination of the radiation from incandescent vapors shows that many spectral lines are emitted by them which lie in the invisible parts of the spectrum. And similarly, multitudes of Fraunhofer lines are detected in the invisible parts of the solar spectrum.

The shortest waves in the extreme violet are about 0.0004 millimeters long; the longest in the extreme red, about 0.0007 millimeters long. Photographic methods have detected waves 0.0001 millimeters long, and observations of the heating effect have detected waves 0.10 millimeters long.

372. **Spectral Series.** — Balmer showed that the spectral lines of hydrogen are so disposed that their wave lengths, or the number of vibrations corresponding, can be calculated by the aid of a general

formula. In applying this formula each spectral line is assigned one of the natural numbers, according to its position in the hydrogen spectrum, beginning with 3, assigned to the line corresponding to the longest wave length, and the insertion in the formula of the number assigned to the line leads to the number of vibrations corresponding to the line. Similar series, calculable by similar though more complicated formulas, have been discovered in the spectra of many other elements.

373. **Distribution of Energy in the Spectrum.** — The heating effect of a limited part of the spectrum has been studied by Langley and others with an instrument invented by Langley, and known as the bolometer. It is simply a narrow and thin strip of blackened iron or platinum, which is so connected in an electric circuit that its resistance can be measured. If the strip absorbs heat, its resistance increases, and the change in the resistance will determine its change of temperature, and so the heat absorbed by it.

With this instrument the heating effect of different parts of the spectrum of a black body has been thoroughly examined. It has been found that the energy E emitted with light of any wave length λ depends upon the wave length and upon the absolute temperature T of the source of the light in a way which has been expressed by Planck in the formula

$$E = \frac{c\lambda^{-5}}{e^{\frac{c^2}{\lambda T}} - 1}.$$

In this formula c is a constant to be found by experiment, and e is the base of the natural logarithms.

The wave length which corresponds to the maximum emission of energy is shorter as the temperature rises. In the light from the sun the maximum energy is emitted with waves whose length is greater than that of the longest waves which give visible red. It has been shown by Wien that the wave length λ_m, corresponding to the maximum emission of energy, and the temperature of a radiating black body are connected by the simple law

$$\lambda_m T = A, \text{ a constant,}$$

and that the energy E_m at the maximum is connected with the temperature by the law

$$E_m T^{-5} = B, \text{ a constant.}$$

374. **Fluorescence.** — Certain substances, such as fluorspar or solutions of chlorophyll or of sulphate of quinine, when placed so

as to receive a narrow beam of light, become self-luminous in a peculiar way. The light emitted by them seems to originate in the path of the beam, and has a characteristic color and spectrum, depending on the nature of the substance. The phenomenon is known as *fluorescence*. According to Stokes the wave length of the emitted light is always less than that of the light which excites it.

By rapidly cutting off and renewing the incident beam, it has been shown that the emitted light persists, though often only for a short time, after the incident beam is cut off. Fluorescence is therefore apparently not distinct from *phosphorescence*, that is, from the phenomenon of persistent luminescence excited by exposure to light in certain substances, such as sulphide of calcium.

Lenard and Klatt have shown that the phosphorescence exhibited by sulphide of calcium and other similar substances is due to the presence of traces of metallic impurities. Lenard explains phosphorescence and fluorescence also as due to vibrations set up within the atoms emitting the light by the return into them of electrons which have been ejected from them under the action of the incident light.

375. Zeeman Effect. — It has recently been discovered by Zeeman that the vibrations of a vapor which emits light are peculiarly modified if they are executed in a magnetic field. We need not describe this modification further than to say that a single spectral line which the vapor would ordinarily emit is divided, when the vapor is in a magnetic field, into two or more lines, and that the light in these lines is differently polarized.

We are not able to explain the Zeeman effect, or indeed to give an adequate explanation of the hypotheses upon which our whole theory of light has been based, by any purely mechanical theory of the ether and of the nature of light. It is now quite certain that what we have called the vibrations of light are periodic electric disturbances in the ether, and that the various modifications impressed upon them by material bodies are due to the electric relations of those bodies. We shall discuss the electromagnetic theory of light in connection with our study of electricity.

376. Effect on the Velocity of Light of the Motion of Bodies. — When the wave theory of light was first studied by Fresnel and Arago, the question of the aberration of light had to be considered (§ 333). Arago perceived that, on the accepted explanation of aberration, the aberration of a star ought to be different from that ordinarily obtained for it, if the tube of the telescope by which it

was observed was filled with water. Observation showed, however, that the aberration was the same when thus determined as it had previously been found to be. It was shown by Fresnel that this result could be explained if the velocity of light was supposed to be affected by the velocity of the body through which it was passing, according to a certain law. In order to test this hypothesis, Fizeau observed the change produced in the velocity of light by sending it through a stream of water. The results of his observations, which have since been confirmed by Michelson and Morley, were in agreement with Fresnel's formula.

These observations are consistent with the hypothesis that the ether is at rest, and does not share in the motion of bodies moving through it. On the other hand, Michelson and Morley, by the use of the interferometer, compared the velocity of light in the direction of the earth's motion with that of light perpendicular to the earth's motion, and found that they could not detect any difference, such as would be expected if the ether is at rest. This result is at variance with most of the other facts known bearing on the subject, and while it is accepted as certain, the conclusion that the ether moves with the earth is not generally drawn.

The question of the action of moving matter upon the ether, and of the way in which the motion of matter affects the progress of light through it, has not yet been solved.

BOOK VI

MAGNETISM AND ELECTRICITY

CHAPTER XVIII

MAGNETISM

377. The Lodestone. — From the earliest antiquity it has been known that certain minerals exist which show a peculiar attraction for iron. Pieces of these substances are called magnets or lodestones. In view of the fact that pieces of iron or steel artificially prepared, and possessing this same property, are also called magnets, it may be best always to designate these natural magnets as *lodestones*. They contain principally magnetic oxide of iron.

The attractive power of the lodestone for iron is shown more strongly at some parts of it than at others. By careful selection, lodestones may be found in which the regions of strongest attraction are two in number, but none are ever found with only one such region.

It was found by the early observers that when a small piece of iron, such as an iron finger ring, was attracted by the lodestone, it also acquired the property of attracting iron. The iron thus attracted by it acquired in its turn the same property of attraction. The attractive force developed in each successive piece of iron decreased in intensity as the iron was further removed from the original lodestone. These pieces of iron are said to be magnetized by induction. Later observations showed that iron can be magnetized by induction when it is brought near the lodestone, or near another magnet, without being in contact with it.

When a lodestone exhibiting two centers of attraction was placed in a light vessel and floated on the surface of water, it always turned about so that one of the two centers pointed toward the north and the other toward the south. This property of assuming a definite direction is generally better exhibited by artificial magnets, in which the two centers of attraction are more precisely developed, than it is by the lodestone.

The compass is a magnet so arranged that it can turn freely in a horizontal plane and indicate by the direction in which it points the north and south line.

378. Gilbert's Study of Magnets. — The first scientific study of magnets was made by Gilbert (1600). He found that when the ends of two lodestones of the simplest type were brought near each other, they were attracted to each other if one of them was an end which would point toward the north and the other an end which would point toward the south. If ends which would point toward the north, or ends which would point toward the south, were brought near each other, they repelled each other. Gilbert called these ends, at which the forces of attraction and repulsion were most strongly exhibited, *poles*, the one which pointed toward the north being called the north pole, the other the south pole. We may express Gilbert's discovery by saying that a north pole of a magnet will attract a south pole of another magnet, and that the north poles or the south poles of two magnets will repel each other.

Gilbert conceived of the magnetic condition of a lodestone, or of any magnet, as due to some sort of arrangement throughout its whole body. The experiment by which he was led to this conclusion consisted in cutting a lodestone exhibiting two poles into two parts, by a section across the lines joining the poles (Fig. 219).

Fig 219.

When the two parts thus made were separated and examined, it was found that the poles which they originally possessed had not been altered, but that two new poles had been developed, one in each piece, of a sort different from that already in it. Each piece was therefore a complete magnet, having north and south poles. When the newly made north and south poles were brought near each other, they attracted each other, and when the two pieces were allowed to meet, the new poles disappeared, and only the original poles remained. It has been shown that whenever a piece of any size is cut off from a magnet, it is always itself a complete magnet, and it is inferred that if the molecules of a magnet could be separated from one another, each of them would be a complete magnet.

379. Artificial Magnets. — As has already been described, a piece of iron may be made a magnet by bringing it near a lodestone. If it is removed from the lodestone and tested, it may remain a magnet still, but the slightest disturbance of it, by striking or

jarring it, will cause it to lose its magnetic condition. If the iron is in the form of steel, however, it will retain the magnetic condition induced in it by the lodestone to a very great degree. A piece of steel thus prepared is called an *artificial magnet*. Since its magnetic condition is retained by it indefinitely, in ordinary circumstances, it is also called a *permanent magnet*. It is with such steel magnets that we can most easily carry out the experiments which Gilbert described as carried out by him with the lodestone.

380. **Magnetization by Induction.** — To study *magnetization by induction* we present one end of a soft iron rod to the north pole of a permanent magnet, and examine the magnetic condition of the rod (Fig. 220). It is found that the rod is magnetized in such a way that the end near the north pole of the magnet is a south pole, and the other end is a north pole. If one end of the rod is presented to the south pole of the magnet, that end becomes a north

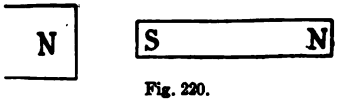


Fig. 220.

Fig. 221.

pole, and the more distant end a south pole. A similar result is obtained if one end of a row of short iron rods is presented to the pole of the magnet (Fig. 221). Each of the rods then becomes a magnet, with a pole different from that of the inducing magnet in the end nearer it, and a pole like that of the inducing magnet in the more distant end.

We shall subsequently describe the region around a magnet, in which magnetic force can be perceived by the aid of another magnet, as a magnetic field, and we shall then describe magnetic induction in a somewhat more general way. For the present it is sufficient to say, that a magnetic pole will induce a pole unlike itself in that part of a piece of iron which is nearest to it. The attraction between these two unlike poles draws the iron and the magnet together. It was this attraction which was first observed, and which was for a long time supposed to be the fundamental property of a magnet. It was not until Gilbert examined the action of one magnet on another, and observed the magnetization by induction, that the true relations of a magnet to iron could be understood.

381. **Making Magnets.** — A piece of steel may be made a magnet by bringing it into contact with a lodestone or a pole of another magnet. Various operations, by which this contact is made according to certain rules, have been devised, in order to effect this magnetization as uniformly and as powerfully as possible. These methods have all been superseded by a method which depends upon the fact that an electric current sets up a magnetic field

around itself. The wire carrying the current is wound into a spiral coil, and the steel bar which is to be magnetized is placed in the axis of the coil. The magnets produced in this way are in every way better than those produced by the older methods of contact.

If two similar bar magnets are placed side by side with their like poles contiguous, they produce a magnet which is more powerful than either of them. By larger combinations of magnets, very powerful permanent magnets can be produced.

382. The Earth as a Magnet. — Attention has already been called to the fact that a magnet which is free to turn in a horizontal plane will point with one of its ends toward the north. The vertical plane containing the direction in which it points at any place is called the *magnetic meridian* for that place. If the magnet is mounted on a horizontal axis so that it can turn freely in the magnetic meridian, its north pole, in the northern hemisphere, will generally point downward, making a certain angle with the horizontal which is called the *magnetic dip* for the place. In the southern hemisphere the north pole points above the horizon. By magnetizing a globe of iron and examining the behavior of small magnets, suspended freely at different points on its surface, Gilbert reproduced on a small scale the phenomena exhibited by free magnets with respect to the earth (Fig. 222). He thus concluded that the earth itself is a magnet, or at least has magnetic poles like those of a magnet, the south magnetic pole of the earth, toward which the north pole of the suspended magnet points, being situated in the northern hemisphere.

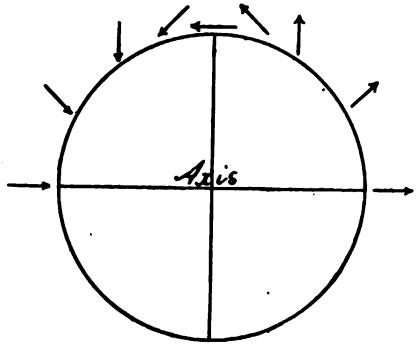


Fig. 222.

This conclusion of Gilbert's is in general correct, but there is still much that is not understood about the magnetic condition of the earth. The magnetic meridians do not coincide with the meridians of longitude, nor do the lines of equal dip coincide with the parallels of latitude. The magnetic equator, at which there is no dip, is an irregular line crossing the geographical equator in two points. The angle between the magnetic meridian and the geographical meridian at any place is called the *magnetic variation* or *declination*

at that place. This variation not only differs in different places, according to no well-defined law, but it also changes with lapse of time. It never exceeds a certain limit, however, and appears after many years to go through a cycle of values. The variation is also subject to a small diurnal change. The dip at any place is also subject to similar changes.

383. Law of Magnetic Force. — Hitherto we have described in general terms the way in which one magnet acts on another. In order to construct instruments for use in magnetic measurements, and in order to advance the theory of magnetism, experiments were carried out by Coulomb to determine the law of the force between two magnetic poles; that is, to determine the way in which the force between the poles depends on the poles themselves and on the distance between them. In these experiments Coulomb used the instrument called the torsion balance. This instrument depends on the use of a twisted wire for measuring the couple by which it is twisted, and is similar in all essential features to that used by Cavendish in determining the gravitation constant (§ 146).

With this torsion balance Coulomb proved, first, that the magnetic forces exerted upon the magnet by the earth are always applied at the same two points in the magnet, which points are the *north* and *south poles*. The line joining these points is called the *magnetic axis*, and when the magnet is free to turn, the magnetic axis indicates the magnetic north and south line. The forces acting on these poles are equal, parallel, and oppositely directed, so that when the magnet is not in the magnetic meridian, it is acted on by a couple which tends to turn it into that meridian. In observations of the action of one magnet on another, this couple due to the earth must be determined and allowed for. When Coulomb determined the forces exerted between two magnetic poles at different distances from each other, he found that these forces were to each other inversely as the squares of the respective distances between the poles.

This conclusion that the force between two magnetic poles varies inversely with the square of the distance between them, was afterwards confirmed by Gauss in another way. Gauss assumed the law of inverse squares, and calculated by means of it the forces which one magnet will exert upon another when they are placed in certain different positions relative to each other. He then examined by experiment the actual forces exerted in these different positions, and found them to be in accord with his calculations. From this result he inferred the truth of the law upon which the calculations were based.

384. Strength of Pole. — Coulomb's law of the force between two magnetic poles recalls the law of gravitation, according to which the force between two masses also varies inversely with the square of the distance. The law of gravitation, however, contains another statement, namely, that the force is proportional to the interacting masses. It is plain that in some way the force between two magnetic poles depends on what we may call the strengths of those poles, and we must now inquire how the strength of a pole can be measured, and how the force between two poles depends on their strengths.

In the case of gravitation, the masses which attract each other are already measured in terms of a unit mass, and, in establishing the law, a direct proof was given that the attraction between two masses is proportional to them both. In the case of magnetism, we have no such independent standard of the strength of a pole, or of what we may call, by analogy, the quantity of magnetism at the pole. We have no evidence that there is such a thing as magnetism. We only know that a piece of steel is in a condition in which it exerts peculiar forces on other similar pieces of steel, and we have no way of measuring that condition except by those forces. We are therefore compelled to measure the *strength of a pole* by the force which it will exert on a standard pole according to some assumed law. The law which we assume is the simplest possible, namely, that the strengths of different poles are to each other as the forces which they will exert on the same pole if placed at equal distances from it. When the strength of pole is thus defined, the law of magnetic force may be enlarged into the statement that the force between two poles is proportional to the product of their strengths and inversely proportional to the square of the distance between them. Owing to the fact that the force between two poles depends on the medium which surrounds them, we assume in this definition that the poles are in a vacuum. The force measured in air is not perceptibly different from that in vacuum. By using m and m' to represent the strengths of the two poles, and r to represent the distance between them, we may express the law relating the force F to the strengths and the distance by the formula $F \propto mm'/r^2$.

385. Unit Magnetic Pole. — Our definition of strength of pole has merely compared one pole with another. It is, however, important, and even necessary for magnetic measurements, to adopt a *unit magnetic pole*, in terms of which all other poles may be meas-

ured. The way in which such a pole can be defined is best seen by examining the formula which expresses the force between two poles. By introducing a factor of proportion k we may change the proportion above into the equation

$$F = k\frac{mm'}{r^2}, \tag{123}$$

which expresses the force between the two poles m and m' when the distance between them is r. Now, in the case of the similar formula expressing the law of gravitation, the two masses, which correspond to the two strengths of pole, are supposed to be known in terms of the unit of mass, and the force and the distance are measured, so that by the substitution of these known quantities in the formula, the constant k is determined as the gravitation constant. In the case of magnetism, it is conceivable that an arbitrarily chosen magnetic pole might be taken as a standard, and other magnetic poles always expressed in terms of it. If this were done, the quantities in the formula would all be known, except the factor k, and so k would be determined as the constant of magnetic force. It is, however, practically impossible to construct a magnet of such a sort that its pole can conveniently be used as a standard, and it is equally impossible to preserve it so that the strength of its poles will not change with use and with lapse of time. We therefore proceed otherwise in determining the unit pole. Instead of arbitrarily choosing a unit pole and then, by observation of the force between two measured poles, determining the value of the factor of proportion, we arbitrarily choose a value of that factor, and then, by observation of the force between two poles, determine their values, or rather the value of their product. To make this choice as simple as possible, we set $k = 1$. We then have,

$$F = \frac{mm'}{r^2}. \tag{124}$$

On this assumption, the measured force between two poles, at a known distance apart, will enable us to determine the product of their strengths. If the poles are alike, we may determine the strength of either one of them. In particular we may define the *unit pole*, or pole of unit strength, as that pole which will repel another equal and similar pole at unit distance with unit force. In the c.g.s. system, the unit magnetic pole is one which will repel an equal and similar pole, at the distance of one centimeter, with the force of one dyne.

386. Magnetic Moment. — In the study of the action of a magnet, we generally find that the formula obtained contains the product of the strength of pole of the magnet and the distance between its poles. This product is called the *magnetic moment* of the magnet. It is constant so long as the magnetic condition of the magnet is unaltered, and is a characteristic of the magnet. Two magnets whose moments are equal are in most respects equivalent as regards their effects in the region around them.

387. Magnetic Field. — When a small magnet is brought near another magnet, it will be acted on by magnetic forces, and will indicate this action by assuming some definite position. If we imagine it possible to isolate one of the poles of this magnet and use it to test the field, or, what amounts to the same thing, if we examine the force exerted on one pole of the magnet, we shall find that, in these circumstances, it is acted on by a force of definite amount and direction. By using such an isolated pole, the forces in the region around a magnet, or around any combination of magnets, can be mapped out with their magnitudes and directions. Such a region, in which a magnetic pole will be acted on by magnetic force, is called a *magnetic field.*

The existence of a magnetic field may often be detected when there is no evidence of the presence of a magnet to which the field can be assigned. Thus there is a magnetic field around a wire carrying an electric current. The earth also carries with it a magnetic field which is not certainly due to the presence of magnetized iron within the earth.

388. Magnetic Intensity or Strength of Field. — If the magnet with which we examine a magnetic field is one whose poles are of unit strength, the force which is exerted on its north pole at any point in the field is called the *magnetic intensity,* or the *strength* of the magnetic field, or simply the *magnetic force* at that point. We designate the intensity by H. The force exerted on any other pole of strength m at the same point is given by $F = mH$.

The field of unit intensity is one in which the unit pole is acted on by unit force. In the c.g.s. system the field of unit intensity is one in which the unit pole is acted on by a force of one dyne. The intensity, from its definition, is plainly a vector, and may be resolved into components, or two intensities may be compounded to find a resultant, by the parallelogram law.

389. Magnetic Field of the Earth. — A great deal of attention has been devoted to the study of the magnetic field of the earth,

partly for scientific reasons, but mainly on account of its great importance in navigation and surveying. The elements of the magnetic field which are examined by experiment are the variation or declination, the dip, and the horizontal intensity, that is, the horizontal component of the strength of field.

The *declination* is determined simply by observing the angle between the geographical meridian and the axis of a magnet suspended so as to turn freely in a horizontal plane. Similarly the *dip* is determined by observing the angle between a horizontal plane and the axis of a magnet suspended so as to swing freely in the vertical plane of the magnetic meridian.

The *horizontal intensity* may be determined either relatively or absolutely.

If a small magnet is suspended so that it can swing freely in the horizontal plane, and is then turned out of the magnetic meridian, it will be acted on by a couple tending to bring it into the magnetic meridian. For a swing of only a few degrees, the moment of this couple is proportional to the angular deviation from the magnetic meridian. The magnet therefore executes oscillations which are similar to those of a pendulum, and the time of oscillation is expressed by a formula like the pendulum formula, in which, however, the moving force is not the weight of the body, but is due to the forces exerted by the field upon the poles of the magnet. The observations enable us to determine the product of the magnetic moment of the magnet and the horizontal intensity; and if the magnet can be protected from alteration, so that its magnetic moment remains constant, observations made at different places on the earth enable us to compare the horizontal intensities at those places.

The value of such determinations as these depends on the constancy of the magnet with which they are made. By using the same magnet to deflect another one from the magnetic meridian, the ratio of the magnetic moment of the magnet to the horizontal intensity can be determined, and by a combination of this relation with the one obtained from the former experiments, a measure of the horizontal intensity can be obtained in absolute units, and independent of the particular magnet with which the experiments are made.

It is found that the horizontal intensity, like the variation and the dip, undergoes both secular and daily changes. Notwithstanding the great amount of study which has been devoted to the earth's magnetism, no satisfactory theory of its origin has yet been given.

EXAMPLES, XX

1. *To find the magnetic intensity at a point on the axis of a short bar magnet.*

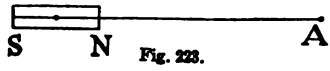


Fig. 223.

Using l to represent half the length of the magnet, and a to represent the distance from the center of the magnet to the point, we have (Fig. 223) the intensity F given by

$$F = \frac{m}{(a-l)^2} - \frac{m}{(a+l)^2}; \quad \text{or} \quad F = \frac{m((a+l)^2-(a-l)^2)}{(a^2-l^2)^2}.$$

If we drop all terms as negligibly small which contain l^2, or higher powers of l, we obtain $F = 4ml/a^3$. Now $2ml = M$, the magnetic moment of the magnet, so that we have finally $F = 2M/a^3$. This force is directed along the axis and away from the magnet or toward it according as the point A is nearer the north or the south pole.

Fig. 224.

2. *To find the magnetic intensity at a point in the equatorial plane of a short bar magnet.*

The forces acting at P (Fig. 224) are $\frac{m}{\overline{PN}^2}$, $-\frac{m}{\overline{PS}^2}$, and may be represented by the equal lines $\overline{PD}$, $\overline{PE}$ respectively. The force desired is the resultant $\overline{PH}$. From the construction $\overline{PH} : \overline{PD} = SN : PN$. Now $SN = 2l$, $PN = a$ approximately, the distance from the center of the magnet to the point P, $\overline{PD} = m/a^2$ approximately, and therefore $F = \overline{PH} = 2ml/a^3 = M/a^3$.

3. *Two short bar magnets are placed at a considerable distance apart, one of them with its axis in the line joining their centers, the other with its axis perpendicular to that line. To find the couples exerted on these magnets.*

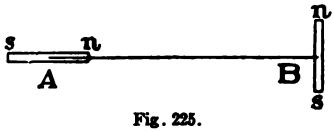


Fig. 225.

a. Magnet A lies in the line joining the centers; magnet B is perpendicular to it (Fig. 225). We shall investigate the approximate value of the couple acting on magnet B.

The magnet B is supposed to be so short that the intensity due to the magnet A at the points occupied by its poles will be approximately equal to that at its center. We use m' to represent the strength of the pole of the magnet B, l' its half length, and so $M' = 2m'l'$, its magnetic moment. The distance between the centers of the magnets is represented by a. Then the force on the north pole of B is (Example 1) $2m'M/a^3$. That on the south pole has the same numerical value and acts in approximately the opposite direction, both forces being approximately parallel to the line joining the centers of the magnets. The moment of the couple acting to turn the magnet B is therefore

$$4m'l'M/a^3 = 2MM'/a^3.$$

b. The couple acting on the magnet A is found approximately, if A can be considered so short that the intensity due to B is the same at each of its ends as at its center, by using the result of Example 2. The force on each of the poles of A is mM'/a^3, and these forces form a couple whose moment is $2mlM'/a^3 = MM'/a^3$.

4. *To find the couple acting on a magnet whose axis makes an angle with the direction of the magnetic force in a uniform field.*

Let the angle made by the axis with the direction of the force of the field be represented by ϕ, and the intensity in the field by H. The force on the north pole of the magnet is $F = mH$. That on the south pole has the same value and is oppositely directed. The arm of the couple (Fig. 226) is $p = 2l \sin \phi$, and the couple is therefore given by $Fp = 2mlH \sin \phi = MH \sin \phi$.

5. *The positions of Gauss.*

If the line joining the magnets A and B in Example 3a is perpendicular to the magnetic meridian, and if the magnet B is pivoted at its center so that it

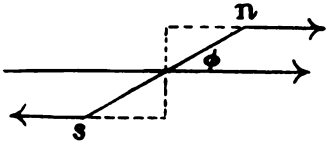


Fig. 226.

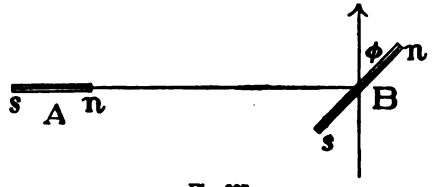


Fig. 227.

can turn in a horizontal plane, the couple due to A will deflect it, and it will assume a position in which it will be in equilibrium between the couple due to A and the opposing couple due to the magnetic field of the earth (Fig. 227). We use H to represent the horizontal intensity of the earth's field, and ϕ to represent the angle of deflection of B from the magnetic meridian. The couple due to the magnet A will have the same forces acting as in the case considered in Example 3a, but the arm of the couple will be $2l' \cos \phi$, so that the couple will be equal to $2MM' \cos \phi / a^3$. The field of the earth will set up a couple (Example 4) equal to $M'H \sin \phi$. When the magnet B is at rest we have therefore $\frac{2MM'}{a^3} \cos \phi = M'H \sin \phi$, or $\frac{M}{H} = \frac{a^3}{2} \tan \phi$.

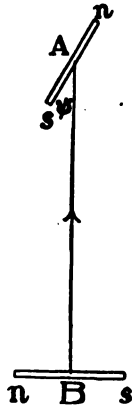


Fig. 228

Observations of a and ϕ enable us to determine the ratio M/H. The magnets in this observation are said to be in the first position of Gauss.

If the line joining the magnets is in the magnetic meridian, and if A is pivoted at its center so that it can turn in a horizontal plane, the couple due to B will deflect it, and it will assume a position in which it will be in equilibrium between the couple due to B and the couple due to the magnetic field of the earth (Fig. 228). Using the result of Example 3b, and setting these couples equal, we get $\frac{MM'}{a^3} \cos \psi = M'H \sin \psi$, or $\frac{M}{H} = a^3 \tan \psi$.

Observations of a and ψ in this case enable us to determine the ratio M/H. By comparison with the previous case we see that for the same distance a we have $\tan \phi = 2 \tan \psi$. The magnets in this position are said to be in the second position of Gauss.

The formulas which give M/H from observations of magnets in the two positions of Gauss were calculated on the hypothesis that the magnetic force due to a pole obeys the law of inverse squares. Presumably the magnetic moment

M and the horizontal intensity H will remain constant during a set of observations, so that if the law of inverse squares is true the relation obtained between $\tan\phi$ and $\tan\psi$ ought to be found to hold by observation. The fact that it does hold is good evidence for the validity of the law of inverse squares.

6. *To find the period of oscillation of a magnet.*

The magnet is suspended so as to turn freely about its center in a horizontal plane. If it is slightly deflected from the magnetic meridian it will oscillate to and fro about its position of equilibrium in the magnetic meridian. The moment of couple which acts on it when it is deflected through the angle ϕ is (Example 4) $MH\sin\phi = MH\phi$, if the deflection is so small that the arc may be substituted for its sine. The angular acceleration α is then given by (§ 100) $I\alpha = -MH\phi$.

As in § 105, we show that the motion is simple harmonic of the period

$$T = 2\pi\sqrt{\frac{I}{MH}}.$$

By observations of the period of an oscillating magnet of known moment of inertia we can determine the product MH.

7. *To find the horizontal intensity of the earth's magnetism.*

By observations of the deflection of an auxiliary magnet B by a magnet A in the first position of Gauss (Example 5 a) we find $\frac{M}{H} = \frac{a^3}{2}\tan\phi$, and by oscillating the magnet A (Example 6) we find $MH = \frac{4\pi^2 I}{T^2}$. We then find H by eliminating M from these equations.

In case we wish to know the magnetic moment M we can find it by making similar observations and eliminating H from the equations.

The value of H is given in terms of quantities which depend only on the fundamental units of mass, length, and time, and is therefore an absolute value.

The Magnetic Field

390. Lines of Force. — If the unit north pole with which a magnetic field is tested is allowed to move in the direction of the force acting on it, it will trace out a continuous curve. This curve is called a *line of force*. It is manifestly such a line that the tangent to it at any point lies in the direction of the force at that point.

These lines of force in magnetic fields due to magnets begin on north poles and end on south poles. Faraday exhibited them by laying a sheet of stiff paper over a magnet and sprinkling iron filings on it. The filings are magnetized by induction so as to form little magnets, with their axes in the direction of the magnetizing force, and so they are attracted to one another, and gather in rows or lines, which mark out the lines of force.

391. Magnetic Potential. — If a unit north pole is moved from one point in a magnetic field to another, there will generally be work done on it by the forces of the field. To obtain an expression for the work that will be done during the motion, we consider the

elementary case in which the field is that set up by a single north pole of strength m. The force which is exerted on the unit pole, when its distance from the central pole is r, is m/r^2 and is directed away from the center. If the unit pole moves to a point at the distance x from the center, it may be shown by a method precisely like that used in Examples, XI, 6, for the analogous case of gravitational force, that the work done on the pole by the force in the field during the motion is measured by

$$m\left(\frac{1}{r}-\frac{1}{x}\right).$$

It is independent of the particular path pursued by the moving pole, and depends only upon the initial and final distances of the pole from the center of force. If the field is due to two or more central poles, the work done will be given by

$$m_1\left(\frac{1}{r_1}-\frac{1}{x_1}\right)+m_2\left(\frac{1}{r_2}-\frac{1}{x_2}\right)+m_3\left(\frac{1}{r_3}-\frac{1}{x_3}\right)+\text{etc.},$$

in which the terms relate to the different poles which together set up the field. Plainly the different values of the distances r and x depend upon the initial and final positions of the moving pole, and do not depend upon the path which it pursues during the motion, so that we may say for this general case that the work done during the motion is independent of the path, and depends only upon the initial and final positions of the unit pole.

The moving pole will not pass beyond the range of the force which acts on it until it is at an infinite distance from all the poles which set up the field. The work that must be done on it to move it to an infinite distance is given by $\frac{m_1}{r_1}+\frac{m_2}{r_2}+\frac{m_3}{r_3}+\text{etc.}=\sum\frac{m}{r}$.

The quantity $\Sigma m/r$ evidently depends upon the strengths and the distribution of the poles which set up the field and is a function of the position of the point determined by the values of r. It is called *the potential* of that point, and is denoted by the symbol V.

If we designate the potential at the point determined by the values of r by V_r, and at the point determined by the values of x by V_x, we may represent the work done on the unit pole as it passes from the one point to the other by $V_r - V_x$. We call $V_r - V_x$ the *difference of potential* between the two points, and may say that the difference of potential between two points in a field of magnetic force is measured by the work done on a unit pole by the forces of the field as it passes from one point to the other.

Similarly, the potential at any point in a field of magnetic force is measured by the work done on a unit pole by the forces of the field as the pole passes from the point to an infinite distance.

392. Equipotential Surfaces. — If we set $V = \sum \frac{m}{r} = \text{constant}$, we may satisfy this condition by continuous variations of the values of r, and thus determine a surface which has the property that at every point of it the potential has the same value. Such a surface is called an *equipotential surface.*

If a unit pole is moved about in an equipotential surface, no work will be done upon it, and therefore the magnetic force of the field can have no component in the surface. The lines of magnetic force are therefore always perpendicular to the equipotential surfaces through which they pass.

If a field is mapped out by means of its lines of force we can describe an equipotential surface in it by beginning at any point and drawing the surface which passes through that point and is everywhere perpendicular to the lines of force. By drawing equipotential surfaces at such distances apart that one unit of work is done on the unit pole as it passes from one of them to the next nearest one, the whole field can be mapped out in a systematic way by its equipotential surfaces.

393. Physical Lines of Force. — In treating the subject of magnetism up to this point, we have followed the course taken by the earlier students of the subject, who were content with the experimental knowledge that forces act between magnetic poles and that they conform to a certain law. We have made no attempt to explain these forces. Faraday thought that it was inconceivable that forces of the nature of magnetic forces could act as they do without the intervention of some medium, through which and by which their action is transmitted. He accordingly adopted what we may call the theory of action through a medium, and undertook to develop the laws of that action, and to discover such properties or conditions of the medium which he assumed as the seat of the action as would account for the magnetic phenomena. His results were afterwards put into mathematical form by Clerk Maxwell, and his general theory, extended to include all the phenomena of magnetism and electricity and the phenomena of light as well, is now generally accepted as affording a satisfactory description or explanation of those phenomena.

In developing this theory Faraday used the term lines of force

with a different signification from that given it in the definition of a previous section. He conceived a magnetic field to be occupied by a medium in a peculiar condition, which he expressed by saying that it contained a certain number of *physical lines of force.* In the field outside of the magnets setting it up, these physical lines of force passed from north to south poles, coinciding in position with the geometrical lines which have already been defined. The physical lines of forces were limited in number, and could hardly be conceived of, except in a very artificial way, unless they were considered as having some thickness, so as to lie side by side of each other, and so to fill the magnetic field. They are more properly called tubes of force, and we shall study them under that name. Faraday's experiments on the magnetic field led him to the statement that magnetic attractions and repulsions are merely consequences of the tendency which the physical lines of magnetic force have to shorten themselves. From the curved shape of these lines he might have concluded, as he did in the analogous case of the electric forces, that the tendency of the lines to shorten themselves, which is equivalent to a tension along the lines, is accompanied by a lateral or transverse pressure equivalent to a dilatation or repulsion of the lines.

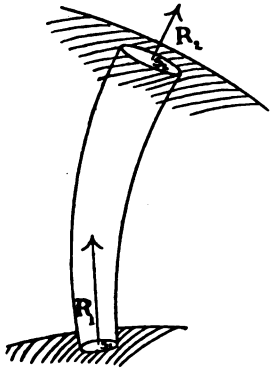
Faraday characterized the strength of a field by the density of distribution of the lines of force in it. This density of distribution is defined by the number of lines of force which pass perpendicularly through a unit of area set up at the point in the field at which the intensity is to be measured, so that we can speak of a strength of field of so many lines. Indeed, this mode of describing the strength of a field is the one which is commonly used in engineering practice.

Of course, to use this mode of representation properly, some convention must be adopted to connect the strength of field and the number of lines, and the relation between the strength at different points in a non-uniform field and the divergence or convergence of the lines must be studied.

394. Tubes of Force. — To do this we proceed to define tubes of force, which are the equivalents of Faraday's physical lines of force, and to study their properties. We consider a magnetic field set up by a distribution of magnets, and to avoid complications we suppose that no other bodies are present in the field, so that the medium surrounding the magnets is not affected by other bodies. Magnetic actions take place in a vacuum, and the medium we assume in this description must therefore be the ether or some

other medium which is not removed when a vacuum is produced. Further study of the magnetic and electric fields will show that there is no need of assuming any other medium than the ether to account for the magnetic and electric actions, and that in fact the properties of the medium which will be deduced from its relations to magnetism and electricity are also sufficient to account for its relations to light and radiant energy in general. The magnetic field in air is so little different from what it would be in vacuum that, for practical purposes, we may consider the field in ordinary space filled with air.

We describe a small area s_1 on an equipotential surface in the field of such a size that the magnetic force R_1 at all points on it has the same value (Fig. 229). The geometrical lines of force which pass through the perimeter of this area are contiguous, and owing to the continuity of the force in the field, remain contiguous to each other in any part of the field in which they are traced. They therefore enclose a region of tubular form, which in general has cross sections of different shape and size in different parts of the field. Such a region is called a *tube of force*. If the surface s_1 is so chosen that we have $R_1 s_1 = 1$, the tube determined by it is called a *unit tube of force*, or simply a *unit tube*. A unit tube thus defined is the equivalent of one of Faraday's physical lines of force.



Figs. 229, 230.

395. Flux of Force. — The product $R_1 s_1$ is called the *flux* of the magnetic force, or the *magnetic flux*, through the area s_1. It is often called the magnetic induction, but as the word induction is used in so many different senses, it is convenient to avoid it if possible.

The flux of force is the same for all parts of a tube of force. This important proposition was demonstrated by Faraday by means of experiments on the production of induced currents by the movements of conductors in a magnetic field. We may prove it otherwise by the help of a geometrical theorem, known as Gauss's theorem. To state this theorem we must first define the flux through a surface a little more generally. The flux through an area of such a size that the force at all points of it has the same value is defined

as the product of the area and of the component of the force normal to the area. Thus, if the total force is R, and if its direction makes the angle α with the normal to the surface in which the area is taken, so that its normal component P is equal to $R \cos \alpha$, the flux through the area s is $Rs \cos \alpha = Ps$. The *total flux* through a closed surface is the sum of all the fluxes through the areas which make up the surface, or $\Sigma Rs \cos \alpha = \Sigma Ps$. In taking this sum the flux will be taken positive or negative, according as the force component is directed outward from the region enclosed by the surface or inward into that region.

Gauss's theorem asserts that in any field of force in which the forces are due to centers attracting or repelling according to the law of inverse squares, the total flux of force taken over a closed surface that does not enclose any of these centers is equal to zero (Examples, XXI, 7).

Let us apply this theorem to the surface formed by the walls of a tube of force and the areas s_1 and s_2 which it cuts out of two equipotential surfaces through which it passes (Fig. 230). The tubular wall contains the lines of force, so that there is no component of force normal to this wall at any point in it, and it contributes nothing to the total flux. The fluxes through the two areas on the equipotential surfaces are $R_1 s_1$ and $R_2 s_2$ respectively, and from Gauss's theorem, $R_1 s_1 - R_2 s_2 = 0$. One of the forces is directed inward through the surface, the other outward, so that in applying Gauss's theorem these products should be taken with opposite sign. We conclude that $R_1 s_1 = R_2 s_2$, and that the product Rs is the same for all parts of the tube.

396. Flux of Force through a Surface Enclosing Magnetic Poles. — In another and more general form, Gauss's theorem asserts that the total flux through a surface enclosing magnetic poles equals $\Sigma 4\pi m$. In this expression m represents the strength of a pole.

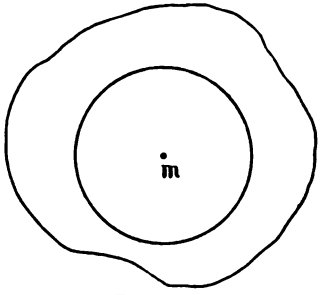


Fig. 231.

We may show this from the first form of Gauss's theorem by considering first the flux from a single pole m (Fig. 231). Around the pole we describe a sphere which is entirely enclosed by the given surface, and we apply Gauss's theorem in its first form to the region between the given surface and this sphere. If P_1 is used to represent the normal component of the force at any point of the given surface, P_2 the normal component at any point of the spherical surface, if s_1

and s_2 are used to represent areas on the given surface and on the sphere respectively, and if we notice that the force is directed inward into the region at one of the surfaces and outward out of the region at the other, we have $\Sigma P_1 s_1 - \Sigma P_2 s_2 = 0$. But the flux through the sphere can be directly calculated; for the force P_2 is the same at every point of it, so that $\Sigma P_2 s_2 = P_2 \Sigma s_2$, $P_2 = \frac{m}{r^2}$, if r is the radius of the sphere, and $\Sigma s_2 = 4\pi r^2$, the surface of the sphere. Hence $\Sigma P_1 s_1 = \Sigma P_2 s_2 = 4\pi m$, or the flux through the given surface equals $4\pi m$.

The same will be true for every pole within the given surface, so that the total flux through a surface enclosing a number of poles is $\Sigma 4\pi m$.

We know from observation that magnetic poles always occur in pairs of equal and opposite strength, so that Σm for a magnet or for any number of magnets is equal to zero, and the total flux through a surface which encloses magnets is equal to zero.

397. Unit Tubes of Force. — Notwithstanding the impossibility of realizing it physically, we have made frequent use of the concept of an isolated magnetic pole. The flux of force through a spherical surface at the center of which such a pole is placed is $4\pi m$. If the surface of the sphere is divided into $4\pi m$ equal areas, since the area of the whole sphere is $4\pi r^2$, each of these areas is equal to r^2/m, and since the force at the surface is equal to m/r^2, the flux of force through each area is equal to 1. The tubes of force which are determined by these areas are therefore the unit tubes defined in § 394, and a magnetic pole of strength m can be considered as associated with $4\pi m$ unit tubes.

The product $Rs = 1$ for all parts of a unit tube. If N represents the number of unit tubes which pass through unit area on an equipotential surface, since s is the cross section of one of these tubes, we have $N = 1/s$, or $R = N$. The intensity in the field at any point is therefore equal to the number of unit tubes which pass through unit area on an equipotential surface at that point.

If the unit tubes are constructed which pass through any equipotential surface in a magnetic field their distribution throughout the field will indicate the intensity in the field at every point. The whole field may therefore be mapped out by means of unit tubes of force.

398. Stresses in the Magnetic Field. — We shall reserve the discussion of the proof that the forces between magnetic poles can be explained by stresses in the medium occupying the field until we have treated the analogous case of the forces in the electric field, in which the discussion is simpler and easier. It will be sufficient here to say that if we assume that the medium in the field is everywhere in a state of stress, defined as a tension along

the lines of force equal to $R/8\pi$, and a pressure at right angles to the lines of force also equal to $R/8\pi$, we can account for the forces between magnets.

399. **Energy in the Magnetic Field.** — The pole of a bar magnet can be represented as a layer or sheet of a hypothetical substance called *magnetism*, distributed uniformly over the end of the bar. The elements of this substance, really the poles of the molecules which constitute the magnet (§ 378), exert magnetic force according to the law of inverse squares.

Let us consider a set of bar magnets, and the field around them. At each pole of strength m there will be a certain potential V, and it may be proved (Examples, XXI, 8) that the work done in constructing the magnets and therefore the energy possessed by them is given by $\frac{1}{2}\Sigma mV$. Now the tubes of force run from north poles to south poles, and a tube which leaves a north pole ends on a south pole of equal strength, so that if we use V_1 and V_2 to represent the potentials at the ends of a tube, the energy is represented by $\frac{1}{2}\Sigma m\,(V_1 - V_2)$, in which m now represents the strength of one of the north poles. If d represents a small distance measured along the axis of a tube of force, Rd represents the work done in moving a unit pole through that distance, and therefore ΣRd, taken over the whole length of the tube, represents the work done in moving a unit pole from one end of the tube to the other. But this work measures the difference of potential, so that we have $\Sigma Rd = V_1 - V_2$. Let us suppose that the energy of the magnets is distributed in the field in such a way that each unit length of each unit tube contains $R/8\pi$ units of energy. Then the length d of a unit tube will contain $Rd/8\pi$ units of energy, and the whole tube will contain $\Sigma Rd/8\pi = (V_1 - V_2)/8\pi$ units of energy. The $4\pi m$ unit tubes which are associated with a pole of strength m will contain $\frac{1}{2}m\,(V_1 - V_2)$ units of energy, and the total energy associated with all the poles will be $\frac{1}{2}\Sigma m\,(V_1 - V_2)$. Now this expression is the same as that which was found directly for the total energy, so that the supposition which has been made as to the mode of distribution of the energy in the field leads to the correct value of the total energy. The energy of the magnets, which they possess in consequence of their relative positions, may therefore be thought of as distributed in the magnetic field, and due to some condition imposed upon the medium occupied by the field.

If a cylinder of unit length is set up on a unit area as base drawn in an equipotential surface, the energy contained in it is equal to that contained in unit length of a unit tube multiplied by the number of

unit tubes which pass through its base. We have shown (§ 397) that the number of unit tubes which pass through unit area on an equipotential surface is equal to R, the magnetic force; so that this product is $R^2/8\pi$. The energy in a magnetic field may be considered as distributed in each unit volume according to this formula.

400. **Magnetic Flux in a Magnet.** — The magnetic flux within a magnet is directed from the south pole to the north pole, and is equal to the magnetic flux outside the magnet, which is directed from the north pole to the south pole. Faraday demonstrated this proposition by experiments on the induced currents set up in circuits properly arranged around a magnet and moved in its field. We may prove it otherwise by the use of Gauss's theorem.

To do so we conceive a bar magnet divided transversely by a very narrow cut, or *crevasse* (Fig. 232). From Gilbert's observations we know that each portion into which the magnet is divided is a complete magnet, and that the faces of the crevasse are therefore equivalent to poles equal and opposite to the original poles in the respective portions to which they belong. In the diagram the face to the left will be a north pole, the face to the right a south pole. We may properly assume that the magnetic distribution within the body of the magnet is not essentially altered by cutting the crevasse.

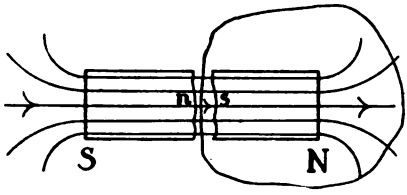


Fig. 232.

We now enclose one of the two portions, say the one containing the original north pole, in a closed surface, which passes through the crevasse. Since the portion within this surface is a complete magnet we shall have $\Sigma Ps = \Sigma 4\pi m = 0$ (§ 396). Now the lines of force outside the magnet all pass from the north pole outward through the surface, so that $(\Sigma Ps)_o$ taken over the portion of the surface which lies outside the crevasse is made up of terms which are all positive and is therefore positive. In order that ΣPs taken over the whole surface shall equal zero, $(\Sigma Ps)_i$ taken over the surface which lies inside the crevasse must be equal numerically to $(\Sigma Ps)_o$ and be negative. The magnetic flux through the magnet is therefore equal to the magnetic flux outside the magnet and is directed in the magnet from its south to its north pole. If the flux is represented as in § 397 by tubes of force, these tubes will be closed tubes, in which the force will be always directed in the same sense along the axis of the tubes. When the force within the magnet is measured in this way,

it is often called the *magnetic induction*, and the tubes of force which represent it are called *tubes of induction*. We may properly call them *tubes of magnetic flux*.

From this point of view a magnet may be considered as a bundle of closed tubes of magnetic flux. The force is directed from south to north pole within the wrapping of the bundle, and from north to south pole outside the wrapping.

EXAMPLES, XXI

1. *To find the relation between the potential and the force at a point.*

Consider two points, A and B, which lie very near together, for which the potentials have the values V_A and V_B. Suppose that the force of the field H has a positive component F in the direction from the point A to the point B, and that the distance between A and B is s. Then the work done on a unit pole as it moves from A to B is given by $Fs = V_A - V_B$;

and hence $$F = \frac{V_A - V_B}{s}.$$

The limit of the ratio $(V_B - V_A)/s$ is the rate at which the potential varies as we proceed through space in the direction from A to B. The component of the magnetic force at the point A in the direction from A to B therefore equals the rate of change of the potential at A with respect to space, taken with the opposite sign.

This property of the potential is sometimes taken for its definition.

The maximum rate of change of potential is evidently in the direction in which the force is greatest, or is along the lines of force. If the force is considered positive in the direction in which a north pole will move, and if the change of distance s is taken positive when drawn in the same direction, the formula shows that for positive values of F and s we have $V_A > V_B$. Motion along the lines of force is from higher to lower potentials.

When the component $F = 0$, $V_A = V_B$, or there is no difference of potential between A and B. The points A and B lie on an equipotential surface and it is perpendicular to the lines of force.

2. *To obtain the potential due to a very short bar magnet.*

Consider a short bar magnet (Fig. 233) with poles of strength m and $-m$ situated at the points N and S respectively. Represent the distance from the center O of the magnet to the distant point P by a, and the half length of the magnet by l. The angle between the line OP and the axis of the magnet is θ.

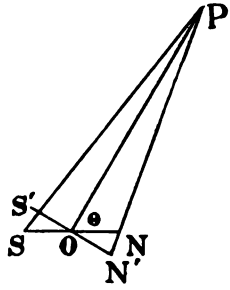


Fig. 233.

The potential at P is $V = \frac{m}{NP} - \frac{m}{SP}$. From the conditions stated NP and SP are approximately parallel, and $NP = a - l\cos\theta$, $SP = a + l\cos\theta$.

With these values we have

$$V = \frac{m}{a - l\cos\theta} - \frac{m}{a + l\cos\theta} = \frac{2ml\cos\theta}{a^2 - l^2\cos^2\theta},$$

or since l is supposed very small, so that l^2 may be neglected in comparison with a^2, we get as an approximate result

$$V = \frac{2ml}{a^2}\cos\theta = \frac{M}{a^2}\cos\theta. \tag{125}$$

3. *To describe the equipotential surfaces around a short bar magnet.*

Choose a certain value, say V_A, for the potential of the surface. Then at a point on the axis (Fig. 234), for which $\cos\theta = 1$, we have $V_A = \frac{M}{A^2}$, in which A is the value of a which is determined by the particular value chosen for V. At points not on the axis, but for which V has the same value, we have $\frac{M}{A^2} = \frac{M}{a^2}\cos\theta$, or $a^2 = A^2\cos\theta$. This is the polar equation of the equipotential surface at which the potential is V_A.

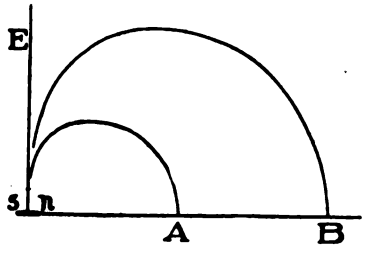


Fig. 234

When $\theta = \pi/2$, $a^2 = 0$, so that the surface passes through the center of the magnet. It will in general be an oval surface passing perpendicularly through the axis at the center of the magnet and at A. A similar surface passing through a more distant point B on the axis will everywhere include the surface just described. The surface of zero potential, for which $V_\infty = 0 = \cos\theta/a^2$, will have $\cos\theta = 0$ for all finite values of a^2, and will therefore be the plane passing perpendicularly through the center of the magnet. It is called the *equatorial plane*. The equatorial plane may independently be seen to be a surface of zero potential, because any point on it is equally distant from the two poles, so that if its distance from either of the poles is represented by d we have $V = \frac{m}{d} - \frac{m}{d} = 0$.

A similar set of equipotential surfaces, in which the potential is negative, may be described so as to include the south pole of the magnet.

4. *To find the potential due to a small disk magnetized uniformly on its faces.*

We conceive the magnetism to be distributed, as if it were a substance, uniformly over the faces of the disk, so that the disk is equivalent to a very short bar magnet with poles of strength m. We use A to represent the area of the disk, and σ a quantity which is called the *surface density* of magnetism, and defined by the equation $\sigma A = m$. Then if the half thickness of the disk is represented by l, we have from Example 2 (Fig. 235),

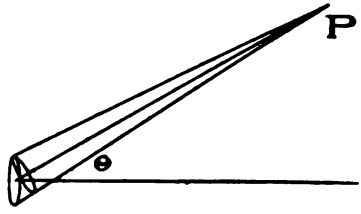


Fig. 235.

$$V = \frac{2ml}{a^2}\cos\theta = \frac{2\sigma lA}{a^2}\cos\theta.$$

Now from inspection of the figure it appears that $A\cos\theta$ is the projection of the area A perpendicular to the line of length a which joins P with the center of the disk; and $\frac{A\cos\theta}{a^2} = \omega$, the solid angle subtended at the point P by this pro-

jection, or by the disk itself. If we further introduce a quantity j, called the *strength* of the magnetized disk, and defined by $j = 2\sigma l$, the potential due to the disk at the point P is represented by $V = j\omega$.

5. *To find the potential due to a uniformly magnetized magnetic shell.*

A magnetic shell is an ideal magnetized body, consisting of a thin sheet of magnetizable matter, magnetized along the normals to its surface. It can be conceived of as made up by setting disks, similar to the one studied in Example 4, edge to edge. If magnetized uniformly the value of j is the same for each of the component disks. In that case the potential at an external point P, due to the shell, is

$$V = \Sigma j\omega = j\Sigma\omega = j\Omega, \tag{126}$$

in which we use Ω to represent the entire solid angle subtended by the edge of the shell from the point P.

6. *To find the work done in carrying a unit pole from a point on one face of a plane magnetic shell to a point on the other face.*

The work done is equal to the difference of potential between the two points. The solid angle subtended by the edge of the plane shell at a point on its face is 2π. The potential on its north face is therefore $2\pi j$, and on its south face $-2\pi j$, so that the work done on a unit pole as it passes from the north to the south face is $2\pi j - (-2\pi j) = 4\pi j$.

7. **Gauss's Theorem.** *To prove that the total magnetic flux through a closed surface which does not enclose magnetic poles is equal to zero.*

Let S (Fig. 236) represent the closed surface and m a magnetic pole to which the force in the field is due. Let s_1 and s_2 represent areas cut out of this surface

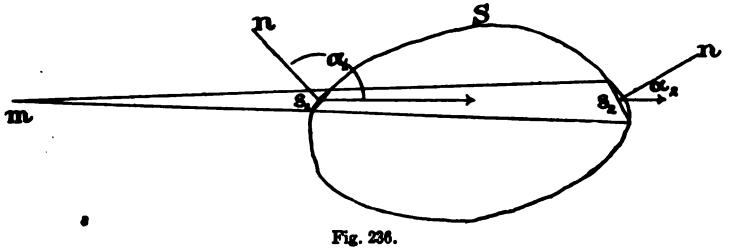


Fig. 236.

by a cone of small aperture so drawn that m is at its apex. Let r_1 and r_2 represent the distances of s_1 and s_2 from m, α_1 and α_2 the angles made with the outwardly drawn normals at s_1 and s_2 by the axis of the cone or the direction of the magnetic force. The magnetic flux through the area s_1 is therefore $ms_1 \cos\alpha_1/r_1^2$, and that through the area s_2 is $ms_2 \cos\alpha_2/r_2^2$. Now $s_1 \cos\alpha_1 = a_1$ is the cross section of the cone at s_1, $s_2 \cos\alpha_2 = a_2$ is its cross section at s_2; and since the cross sections of the cone at different distances from the apex are proportional to the squares of those distances, $a_1 : a_2 = r_1^2 : r_2^2$, or $a_1 = a_2 r_1^2/r_2^2$. Substituting this value of $a_1 = s_1 \cos\alpha_1$ in the expression for the flux through s_1, we find that it is equal numerically to the flux through s_2.

The flux through s_1 is of opposite sign to that through s_2, for the normal at s_1 makes an obtuse angle with the direction of the force, while that at s_2 makes an acute angle with the same direction. The two surfaces therefore contribute equal and opposite terms to the flux through the surface.

The whole surface can be divided by similar cones into similar pairs of surfaces, and since the flux through each pair equals zero, the total flux will also equal zero.

Since any magnetic field can be reproduced by the proper arrangement of magnetic poles, and since the flux of force through a closed surface due to each of these poles is equal to zero, the same will be true for the flux due to an assemblage of poles, and so for the flux in any magnetic field.

If the surface is so folded that the lines of force pass through it more than twice, they will pass through it always an even number of times, and to each pair of such intersections the foregoing demonstration will apply, so that Gauss's theorem holds for a surface of any degree of complexity.

8. *To obtain an expression for the energy of a set of magnets.*

The magnets may be represented by a distribution of magnetism on the surfaces at which their poles lie. Let V represent the potential at any one of these surfaces, and m the strength of the corresponding pole. The poles may be thought of as built up by bringing up elementary quantities of magnetism to each place where a pole is situated; and the work done in building up the poles will be equal to the energy of the set of magnets. Let n represent a large number and m/n the elementary quantity of magnetism brought up at any one time to the place where a pole is situated. The potential at the place is at first 0. When the first element is brought up the potential becomes V/n, and no work is done in bringing up the element. When the next element is brought up, the potential becomes $2V/n$, and work equal to mV/n^2 is done on the element. Similarly for the next element, the work done in bringing it up is $2mV/n^2$; and so on. When the final potential V is reached, the total work done in forming the pole m is

$$\frac{mV}{n^2}(1+2+\ \ldots\ (n-1)) = \frac{mV\,(n-1)\,n}{2n^2} = \frac{1}{2}mV\left(1-\frac{1}{n}\right).$$

If n is very great, or infinite, this reduces to $\frac{1}{2}mV$, and a similar expression holds for the work done in forming each pole. The total energy of the system of magnets is therefore $\frac{1}{2}\Sigma mV$, where the summation is taken over all the poles.

Magnetic Induction

401. Magnetization by Induction. — When a piece of soft iron is brought into a magnetic field it becomes magnetized by induction. The distribution of magnetism in it is described as being along the lines of force; that is, the piece is magnetized as if its north and south poles were free and moved in it in the directions in which such poles would move along the lines of force. This is consistent with the earlier description of induction, according to which a south pole is formed in the iron on the side nearest the north pole of the permanent magnet which sets up the field.

402. Intensity of Magnetization.—The extent to which the iron is magnetized is defined in terms of a quantity called the *intensity of magnetization*. This is defined as the quotient of the magnetic

moment by the volume, or the magnetic moment of unit of volume. It is easily seen that, if the unit of volume is taken small enough, this quantity characterizes the magnetic condition of each part of the iron. The definition assumes that the magnetism exhibited by a magnet is due to a condition existing in every part of it. This assumption is suggested by the experiment of Gilbert already referred to (§ 378) and is supported by other peculiarities of magnets or magnetized iron which will soon be considered.

The intensity of magnetization may be connected with the magnetic poles of a magnet, whether permanent or induced, in the following way: We consider a short bar of length d and of uniform cross section a, uniformly magnetized in the direction of its length. If I represents its intensity of magnetization, its magnetic moment is represented by Iad, and hence, by the original definition of the moment of a magnet (§ 386) Ia is a quantity which is equivalent to the magnetic pole of the bar.

It is often convenient for purposes of calculation to replace the poles of a magnet by distributions of two substances, called *north magnetism* and *south magnetism*. If we use this mode of representation, the poles of a uniformly magnetized bar magnet, in the case before us, may be represented by a distribution over its ends, with a uniform density σ, such that the strength of the pole equals σa. The density σ, in this mode of representation, is equivalent to the intensity of magnetization I.

403. Paramagnetism and Diamagnetism. — Hitherto we have described and studied magnetism as if it were a property only of iron. This, however, is not the case. Iron excels all other bodies, in a very marked degree, in the extent to which it exhibits magnetic properties, and it is apparently the only substance from which permanent magnets can be made. The other metals of the iron group, and especially nickel, are also capable of being magnetized by induction in the way in which iron is. They are called *paramagnetic* bodies. Alloys of copper, aluminium, and manganese have been prepared by Heussler, which exhibit paramagnetic properties almost as marked as those of iron. A few other substances, among them oxygen, either in the gaseous state or in the liquid state, are paramagnetic.

Until 1846 it was supposed that most substances were not affected in the least in a magnetic field, but in that year Faraday, by the use of a very powerful electromagnet, found that all substances upon which he experimented were more or less affected by the magnetic field, but in a different way from that in which the paramagnetic bodies are affected. The effect discovered by him is exhibited most strikingly by bismuth. The contrast between the behavior of bismuth and of iron is shown by the following experi-

ments: If a rod of iron is suspended between the poles of a strong magnet, it will turn and point along the lines of force. The action is in accordance with the general law of induction, which has already been stated. On the other hand, if a rod of bismuth is suspended in the same field, it will turn until it points across the lines of force. Since the bismuth does not retain any traces of magnetization when it is removed from the field, we are not able to tell exactly what condition it assumes when in the field, but its behavior in the field can be explained by supposing that it becomes a magnet, so developed that the positive direction of its axis is opposite to the positive direction of the lines of force of the field. Bodies which exhibit this property are called *diamagnetic* bodies. The action, even in the most pronounced cases, is extremely feeble.

The hypothesis made in the last paragraph, that the magnetic arrangement in a diamagnetic body is opposite to that which is set up in a paramagnetic body, is extremely difficult to reconcile with our theories of magnetism. An experiment tried by Faraday indicates a way by which that difficulty may be avoided. Faraday filled a small glass tube with a solution of sulphate of iron, which is a paramagnetic body. When this tube was suspended in the magnetic field, it pointed along the lines of force. When, however, it was immersed in a stronger solution of sulphate of iron, it pointed across the lines of force. This experiment indicates the following hypothesis to account for the contrast between paramagnetic and diamagnetic bodies; that the ether is itself a magnetizable body, and that paramagnetic bodies are more strongly magnetized than the ether around them, while diamagnetic bodies are less strongly magnetized. There is no way by which these hypotheses can be tested by experiment.

Faraday was led to search for an action of the magnetic field on substances of all sorts by his discovery of an analogous action of the electric field, and he interpreted rightly his discovery of diamagnetism as evidence in favor of his general theory of medium action.

404. Magnetic Force in General. — We have so far studied magnetic forces only in vacuum. When we investigate the magnetic forces within bodies which are magnetized, either as paramagnets or as diamagnets, in the magnetic field, we find that some of our previous statements require modification.

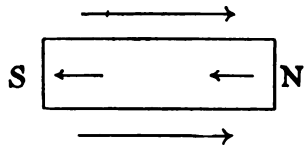


Fig. 237.

Let us consider the force within a small cylinder of a paramagnetic substance placed with its axis coinciding with a line of force in the original magnetic field (Fig. 237). The cylinder will become a magnet, and the field will be modified by its presence. If we consider the magnet thus formed as adequately represented by

two equal quantities of north and south magnetism placed at the ends of the cylinder, and conceive the material of the cylinder removed, the magnetic force at any point within the region occupied by the cylinder is the resultant of the original force of the field and the force arising from the magnetic distribution which replaces the cylinder. This force is called *the magnetic force* within the cylinder and is represented by H. It may be measured ideally without removing the material of the cylinder by cutting out of the cylinder a long cavity, with its axis coincident with the lines of force and of infinitesimal cross section. The magnetic force at a point within this cavity is the resultant of the original force of the field, of the force arising from the magnetic distribution on the ends of the cylinder, and of the magnetic distribution on the ends of the cavity, which arises from the fact that the elements which constitute the body are complete magnets, with their north poles all pointing along the lines of force. This last force is negligible in this case, because the ends of the cavity are of infinitesimal area, and are at finite distances from the point at which the force is measured; so that the force actually measured in such a cavity is the same as that defined before. The definition thus given is called by Lord Kelvin the *polar* definition of magnetic force.

In the case chosen for illustration the magnetic force within the cylinder of paramagnetic substance is manifestly less than the force of the original field at the same point, for the force due to the magnetic distributions on the ends of the cylinder is opposed in direction to the force of the original field. We conclude in general that the magnetic force at a point in a field is diminished by the presence of paramagnetic substance around that point, and is increased by the presence of diamagnetic substance.

The normal component of the magnetic force changes its value abruptly, as we pass along a line of force in the field from vacuum into a paramagnetic or diamagnetic substance, or generally from one substance to another. The tubes of force by which the strength of field is represented will therefore change in number abruptly as they pass from one substance to another.

405. Magnetic Induction. — When we studied magnetic flux (§ 400), we measured the magnetic force in a magnet within a narrow crevasse, cut through the magnet so that its faces were perpendicular to the lines of force. Let us examine the magnetic force measured in such a crevasse within a paramagnetic body magnetized by induction. As before, we consider a short cylinder, placed with

its axis coinciding with a line of force of the original magnetic field. We suppose it cut in two by a crevasse (Fig. 238) and examine the force within the crevasse. If the cylinder is small, and the original field of uniform intensity, we may assume that the cylinder is uniformly magnetized by induction. If I represents its intensity of magnetization and a the area of one of its ends, its strength of pole, or the magnetic distribution at one of its ends, will be represented by Ia (§ 402). The total flux through the cross section of the cylinder or through the faces of the crevasse due to the induced magnetization will therefore be $4\pi Ia$ (§ 396); and since we suppose that the cylinder is uniformly magnetized, so that the magnetic force is the same at all points of the cross section, we obtain for the magnetic force at any point in the crevasse the value $4\pi I$. This force is directed along the lines of force of the original field.

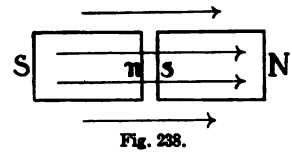


Fig. 238.

The total force in the crevasse is the resultant of the magnetic force H, due to the original field and to the induced poles, and of this additional force, $4\pi I$, due to the distributions of magnetism on the faces of the crevasse. The force measured in this way is called the *magnetic induction*, and is represented by

$$B = H + 4\pi I. \tag{127}$$

The definition thus given is called by Lord Kelvin the *electromagnetic* definition of magnetic force.

The normal component of the magnetic induction changes continuously and not abruptly as we pass along a line of force from vacuum into a paramagnetic or diamagnetic substance, or generally from one substance to another. The tubes of force by which the strength of field is represented will not change in number as they pass from one substance to another, if the strength of field is measured by the magnetic induction. The tubes thus drawn are more properly called *tubes of magnetic induction*, or *tubes of magnetic flux*.

406. **Magnetic Susceptibility.** — Experiment has shown that the intensity of magnetization of a body brought into a magnetic field depends upon the strength of the field. In case the induced magnetism developed is feeble, the strength of the field will be approximately the magnetic force H. It is found that for most substances the intensity of magnetization is proportional to the magnetic force, so that $I = kH$. The factor of proportion k is called the

magnetic susceptibility. It is positive for paramagnetic substances, negative for diamagnetic substances.

407. Magnetic Permeability. — For most substances, therefore, we have

$$B = H + 4\pi I = H(1 + 4\pi k) = \mu H \qquad (128)$$

if we represent $1 + 4\pi k$ by the symbol μ. The factor μ is called the *magnetic permeability*. It is equal to 1 for vacuum, is greater than 1 for paramagnetic substances, and less than 1 for diamagnetic substances.

In the case of a strongly paramagnetic substance like iron, the magnetic permeability is not a constant, but varies with the strength of the field. It can always be defined, and measured, by the ratio of the magnetic induction to the magnetic force.

408. Causes Affecting Magnetization. — If the strength of a magnetic field is gradually increased from an initial value zero, a piece of soft iron placed in the field is magnetized by induction, and its intensity of magnetization increases as the strength of field increases (Fig. 239). The two quantities at first increase proportionally, but as the strength of the field increases, the intensity of magnetization for a time increases faster and then more slowly than the strength of the field, and at last reaches, or nearly reaches, a limit. The iron is then said to be *saturated*. If the strength of field is now diminished until it gradually becomes zero, the intensity of magnetization diminishes, but does not entirely vanish. In order to reduce it to zero the field must be reversed, and its strength in the reverse direction raised to a certain value. After it has passed this value, the intensity of magnetization also reverses, and increases in the reverse direction until the iron is again saturated. By a repetition of this process, the magnetization of the iron can be carried through a cycle of values. For each value of the strength of field there will be two values of the intensity of magnetization, depending on whether the particular strength of field has been reached by increasing, or by decreasing, the field strength. This general behavior of iron, by which its magnetization seems to lag behind the magnetizing force, is called by Ewing, who investigated it, *hysteresis*.

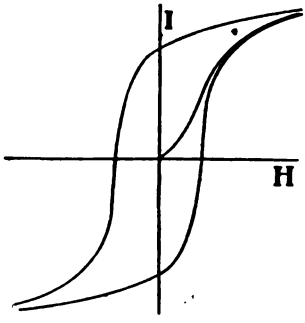


Fig. 239.

If a piece of steel is placed in a magnetic field, it will also become a magnet, though not so rapidly as an equal piece of iron, nor will it attain so high an intensity of magnetization. Its magnetization can be enhanced by striking it so as to make it ring. When it is removed from the magnetic field, it does not lose its magnetism, when struck or jarred, as the iron does, but retains at least the greater part of it. It is thus a permanent magnet. By striking it again, its magnetization will be diminished. If it is heated, its magnetization will also diminish, but will return to its original value, or nearly so, when the magnet is cooled, unless the heating has been carried beyond a certain limit. If the temperature of the magnet is carried above 785° C., the magnetization entirely disappears, and is not restored on cooling. A piece of iron raised above that temperature and brought into a magnetic field exhibits no magnetic properties whatever, until its temperature falls below that temperature.

409. Theories of Magnetism. — When the actions of magnets were first studied, they were ascribed to the presence in the magnet of certain effluvia, or magnetic fluids, which were assumed to be of two opposite kinds, possessing the properties of repelling fluid of the same kind and of attracting fluid of the opposite kind. These fluids were supposed to be present in equal quantities in any mass of iron, and to be separated by the process of induction. The fluid collected in one end of the magnet formed the north pole, and that in the other end, the south pole. Gilbert's discovery that when a magnet was broken the two parts did not contain simply a north and a south pole respectively, but became complete magnets, made it necessary to suppose that these fluids existed in each particle of iron, and were separated in it by induction, so that each particle in the magnetized iron became a magnet. The theory of magnetism was developed in this form by Poisson, and showed itself capable of accounting for the laws of magnetic force, the phenomena of magnetic induction, and of the distribution of magnetic force in a magnet.

This theory does not explain so readily the facts which have been described in the last section. In order to explain them, Weber proposed another theory, which, as modified by Ewing, is the one now generally accepted. It does not attempt to account for magnetism itself, but only to explain the behavior of magnetized bodies. It is rather a theory of the structure of a magnet than a theory of magnetism. Weber supposed each molecule of iron to be a magnet.

In the unmagnetized condition of the iron these molecular magnets have no definite arrangement. The magnetization of the iron consists in arranging the molecular magnets in chains or rows. If we consider the magnetic forces which a single row of molecular magnets thus arranged will exhibit, it appears that the contiguous poles of two neighboring molecules will mutually destroy each other's effects, so that the only poles which will be effective in producing an external magnetic field will be the north pole of a molecule at one end of the row and the south pole of a molecule at the other end. If we consider a magnet to be a collection of such rows laid side by side, and take into account the fact that the mutual repulsion of the similar free poles at the ends of the rows will tend to make those ends diverge from each other, we can account for the existence of magnetic poles, for the distribution of magnetic force over the ends and sides of a magnetized bar, and for the fact that the intensity of magnetization of a magnet is greater in the middle than it is at the ends. The effect produced by striking a magnet is also explained by this theory; for we may suppose that any one molecule of the iron is restrained somewhat from pointing in the direction of the lines of force of the field by the mechanical obstruction of the molecules around it, and that if it is momentarily freed from this obstruction by the jar produced by striking the magnet, it will swing into position more readily; or if it is already in position, and the magnet is not in a magnetic field, will swing more or less out of position, so as partly to disintegrate the rows. The effect produced by heating the magnet, which gives its molecules more freedom of motion, is explained in a similar manner.

To explain hysteresis we must consider, with Ewing, that the arrangement of an assemblage of molecular magnets is not an entirely unordered or irregular one, but that by their mutual attractions the molecules will arrange themselves in stable groups. When a magnetizing force is applied to a collection of such groups, the first effect will be to turn the molecules of the group toward the line of the force, and so to increase the intensity of magnetization. When the magnetizing force reaches a certain limit, the groups, one by one, cease to be stable. The molecules in them then turn freely under the magnetizing force, and form new stable groups, of which the intensity of magnetization is much greater. This description is in accord with the experimental fact that the magnetization of iron undergoes a great increase for a comparatively small change of

the magnetizing force, between certain limiting values. After these new groups are formed, the further increase of the magnetizing force merely brings the molecules of the groups more nearly parallel to one another. Complete saturation will be reached when they are all arranged in parallel rows. When the magnetizing force is again diminished, the stability of the groups that have been formed continues even after the value of the magnetizing force is less than that at which the groups were formed, and the groups do not break down and assume their original condition until the magnetizing force is considerably diminished.

It is evident that this theory of Ewing explains all the facts which we have explained by the theory of Weber.

The discovery by Heussler that alloys containing 60 to 70 per cent of copper, and 40 to 30 per cent of aluminium and manganese, preferably in amounts proportional to their atomic weights, are as strongly paramagnetic as cast iron, opens up the whole question of the structure of a magnet. Copper and aluminium are both diamagnetic substances, and manganese is only feebly, if at all, paramagnetic. No adequate theory has yet been proposed to account for the alloy composed of such substances exhibiting so powerful a magnetic action.

EXAMPLES, XXII

1. *To show how the force in a magnetic field is affected if the field is filled with a substance with magnetic permeability* μ.

The tubes of force of the magnets setting up the field are not altered by the presence of the substance, if the force which determines their distribution is the magnetic induction (§ 405). The magnetic induction B is therefore the same everywhere as the magnetic force of the field before the substance was introduced. But $B = \mu H$, so that H, the magnetic force in the substance, $= B/\mu$, or the original magnetic force is altered in the ratio of 1 to μ by the introduction of the substance.

In the particular case in which the field is due to the presence of a single pole of strength m, the magnetic force at a distance r from it becomes $m/\mu r^2$, and the force between this pole and another one of strength m' is $mm'/\mu r^2$.

2. *To find the distribution of energy in a magnetic field in a substance of magnetic permeability* μ.

The energy of the magnets is still expressed by $\frac{1}{2}\Sigma mV$ (Examples, XXI, 8), since the demonstration of this formula is not affected by the fact that the magnets are surrounded by another medium than ether. In carrying out the demonstration of § 399 the force to be used is $H = R/\mu$, and $\Sigma Hd = \Sigma Rd/\mu = V_1 - V_2$. The energy must therefore be distributed so that each unit length of each tube of magnetic flux shall contain $R/\mu 8\pi$ units of energy.

In unit volume there will be $R^2/\mu 8\pi$ units of energy.

3. *To find the change of energy produced by the introduction into a magnetic field of a small body of magnetic permeability* μ.

We assume that μ is so nearly equal to 1 that the force in the field is not appreciably changed by the introduction of the body. Before the body is brought into the field the energy in the space afterwards occupied by it is equal to its volume multiplied by $R^2/8\pi$. When the body occupies the space, the energy in it is equal to its volume multiplied by $R^2/\mu 8\pi$. The gain in energy by the

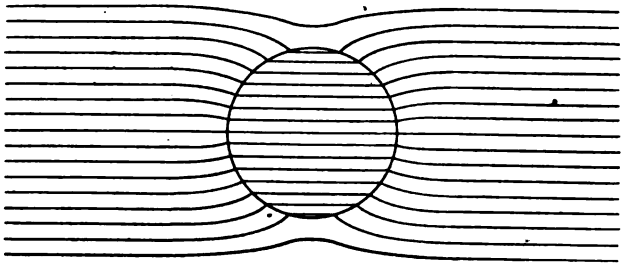


Fig. 240.

introduction of the body is $R^2(1-\mu)/\mu 8\pi$, and is positive or negative according as $\mu < 1$ or > 1.

4. *To show that a paramagnetic body tends to move from a weaker to a stronger part of the field.*

The body tends to move so as to diminish the potential energy of the system. If $\mu > 1$, $R^2(1-\mu)/\mu 8\pi$ is negative and the potential energy diminishes as R^2

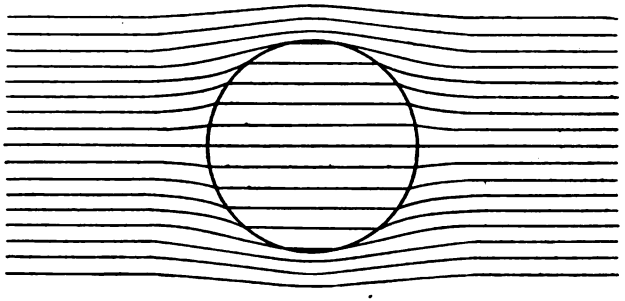


Fig. 241.

increases. The motion will therefore be such as to make R^2 as great as possible. It will be in the direction in which R^2 increases most rapidly, and not necessarily along lines of force.

In the case of a diamagnetic body, $\mu < 1$, and the potential energy diminishes as R^2 diminishes. The motion will therefore be from a stronger to a weaker part of the field.

5. *To show that if a paramagnetic body is introduced into a magnetic field the space which it fills will be traversed by more tubes of flux than before.*

The force R of Example 2 is the same as the magnetic induction B, and the demonstration there given permits us to assert that each unit of length of a

tube of flux contains $B/\mu 8\pi$ units of energy, in which μ represents the particular magnetic permeability of the substance occupying the volume considered. If a body is introduced into a part of the field in which the permeability was equal to 1, and if $\mu > 1$, the energy of the tubes of flux which lie in the body will be less than the energy of the unoccupied ether, and hence the tubes of flux around the body will move into it, so that as many of them as possible lie in the body. Equilibrium is reached when the lengthening of the tubes outside the body increases the energy as much as the entrance of more tubes into the body diminishes it (Fig. 240).

In the case of a diamagnetic body the energy of the tubes of flux which lie in the body will be greater than the energy of the unoccupied ether, and hence the tubes of flux in the body will move out of it, until the lengthening of the tubes outside the body increases the energy as much as the moving of more tubes out of the body diminishes it (Fig. 241).

CHAPTER XIX

STATIC ELECTRICITY

410. Electric Attraction. — If a piece of amber is gently rubbed on a dry cloth, and is then brought near small bits of straw or paper, these light bodies will be attracted by it, and will adhere to it. This observation is a very ancient one. It is said to have been known to Thales (600 B.C.). By the ancient as well as by the mediæval philosophers, this attraction was looked on as something occult, and was supposed to be peculiar to amber, just as the attraction of iron was peculiar to the lodestone. It was not investigated in a scientific way until it was studied by Gilbert (1600) in connection with his study of the magnet. Gilbert mounted a light pointer on a needle, so that it could turn about freely, and with it as an indicator he examined other bodies besides amber, to see whether they would exhibit a similar attractive power. Sulphur had apparently been found by some previous observer to behave like amber, and Gilbert found that many other bodies behaved in a similar way. From the Greek name for amber, *electron*, he called this action *electric action*. We now use the word *electricity* to express the assumed cause of this action. A body exhibiting electric action is said to be electrified, or to be charged with electricity.

Von Guericke (1672) made an important observation by noticing that the light bodies, which were attracted to the electrified body, were repelled by it after a time.

411. Electric Conductors. — In experimenting on electric action, Grey (1731) discovered that when certain bodies were brought into contact with an electrified body they also acquired the power of attracting light bodies. He next undertook to transmit this power from the electrified body to another body through a long thread. In one of his first experiments, he hung an ivory ball by a thread from an upper window, and found that when the electrified body was touched to the upper end of the thread the ball became electrified. He tried next a similar experiment, in which the thread was suspended horizontally by loops of thread, and found that in this case the ball was not electrified. Considering the reason of his success in the one case and his failure in the other, he conceived that it

might be due to the conducting away of the electricity by the loops on which the thread was suspended in the second experiment. He accordingly suspended the thread by other bodies, and found that when silk thread was used for the suspension the action was transmitted to the ball. Grey thus proved the distinction between conductors and nonconductors. Many bodies, of which the metals are the most conspicuous, transmit the electric condition from one body to another, and are classed as *conductors*. Other bodies, of which glass, the resins, and silk are examples, transmit this condition either very slowly, or not to any perceptible degree, and are classed as *nonconductors*. Subsequent investigation has shown that there is no sharp line of separation between these two classes of bodies. Probably all bodies conduct to some extent. The distinction, however, between conductors and nonconductors, is generally made, and many substances are so nearly nonconductors that they can be treated as such in most experiments.

412. Electric Attractions and Repulsions. — Putting together the transmission of the electric condition, as observed by Grey, and the repulsion of light bodies from the electrified amber after contact with it, as observed by von Guericke, it was easy to see that the repulsion might be due to the similar electric condition of the bodies. The experiment was therefore suggested of examining the way in which various electrified bodies act on one another. Such experiments were carried out by Dufay (1733), and a few years later by Franklin. These investigators found that all bodies which can be electrified by friction can be placed in one of two classes, according as they repel or attract an electrified piece of glass.

These experiments are most easily conducted by using as an indicator a gilded pith ball hung by a nonconducting silk thread. If a rod of glass is electrified by friction, and brought near this ball, it will at first attract the ball to it. As soon, however, as the ball acquires the electric condition by contact with the glass, it is repelled from the glass. Its electric condition is then the same as that of the glass, and it can be used instead of a piece of glass in testing other bodies. According to Dufay, it is vitreously electrified, or charged with vitreous electricity. According to Franklin, who adopted another theory and considered the effects to be described as due to the excess or defect in the body of a single substance, it is positively electrified.

To test the electric relations of various bodies, they are submitted to friction, and then presented to the positively electrified indicator. Some of the bodies thus treated will repel the ball as the glass did, and are therefore in the same electric condition as that of the glass. The other bodies will attract the ball. The question arises whether this attraction is similar to that exerted by the glass on the ball in the first instance, or whether it is something different. To test

this, the ball is touched with the hand, so that its electric condition is removed. One of the bodies which causes attraction of the positively electrified ball is then presented to it. The ball behaves exactly as it did when the glass was brought near it; that is, it is first attracted, until it comes in contact with the body, and is then repelled. It is therefore presumably in the same electric condition as that of the body which it has touched. In this condition it is repelled by all the bodies which attracted it when it was positively electrified, and is attracted by all the bodies which then repelled it. The ball has therefore acquired another condition, in which the forces which it exerts are opposite to those which it exerts when positively electrified. This other condition is developed on resin, and Dufay therefore said that the ball in this condition is resinously electrified, or is charged with resinous electricity. In accordance with his theory, Franklin said that it is negatively electrified.

Franklin's terms are the ones which we now use to express these conditions.

Dufay summed up the results of his investigation in the law that there exist two electric conditions, such that two bodies in the same condition repel each other, while two bodies in different conditions attract each other. This law may be otherwise stated by saying that similarly electrified bodies repel each other, while dissimilarly electrified bodies attract each other. As we shall subsequently find that the superposition of dissimilar electricities on the same body results in the disappearance of both of them, we may also speak of oppositely, instead of dissimilarly, electrified bodies.

413. **The Electric Spark.** — Von Guericke constructed an electric machine by mounting a ball of sulphur so that it could be turned around an axis, and setting near it a metallic body supported by a nonconductor. When the ball was turned, and the dry hands, or a piece of flannel, were pressed upon it, it became negatively electrified and imparted negative electricity to the metallic conductor. With this machine, von Guericke found that, after the conductor had been receiving electricity for some time, a spark would pass from it to any conducting body held near it.

Other electric machines were soon constructed which were far more efficient than this primitive one, and with them much larger sparks were obtained. These sparks were generally taken to be evidence of the passing of something, called electricity, from the charged body.

414. **The Leyden Jar.** — Cunæus of Leyden (1745) attempted to collect electricity in water by holding a glass of water near the conductor of an electric machine, and allowing sparks to pass through a nail partly immersed in the water. After the sparks had passed for some time he attempted to withdraw the nail, and in so doing received a shock through his arms and body. In the discovery

of this effect he had been anticipated by von Kleist, but Cunæus's account of his discovery attracted attention, and the instrument by which the shock could be given was called the Leyden jar. This name is still in use, although the Leyden jar is simply one example of the class of instruments known as electric condensers.

In its ordinary form the *Leyden jar* is a glass jar coated within and without, over its bottom and the lower part of its sides, with tinfoil. Through the stopper in the neck is fixed a metallic rod, ending above in a ball or knob, and connected with the inner coating of the jar by a chain. To charge the jar the outer coating is touched with the hand, or is otherwise brought in conducting contact with the earth, and sparks are allowed to pass to the knob. After they have passed for some time, the jar may be discharged by touching one end of a conducting body to the outer coating, and bringing the other end of it near the knob. A spark will pass between the knob and the conductor, which is brighter and thicker and makes a louder report than any of the sparks by which the jar was charged. The successive charges which pass to the jar from the machine seem to have been accumulated or condensed in the jar. When the discharge is taken through the body, the shock is often very severe.

The explanation of the action of the Leyden jar was first given by Franklin (1747). In order to appreciate it, we must consider the way in which an electrified body acts on bodies near it, to produce the electric condition in them.

415. Electric Induction. — In examining the effects of an electrified body on other bodies, it is necessary to mount these bodies on nonconductors, or to *insulate* them. When a conductor thus insulated is brought near an electrified body, it will exhibit signs of electrification. It is said to be charged by induction.

We test the peculiarities of its electrification by the use of a proof-plane, a small disk of sheet metal mounted on a nonconducting handle. When the proof-plane is laid on the surface of the conductor at any point, it acquires a charge similar to that of the conductor at that point, and by removing it and testing the charge on it the charge on the conductor can be determined. By the use of such a proof-plane it is shown that the parts of a conductor charged by induction which are farthest from the electrified body (Fig. 242) have a charge like that of the electrified body, and that those parts which are nearest the electrified body have a charge of the opposite kind. These charges are shown to be equal in amount by removing

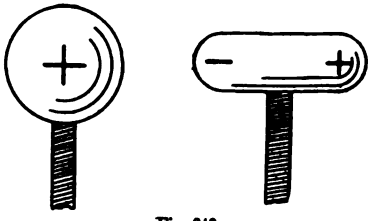


Fig. 242.

the conductor from the presence of the electrified body, for then no charge can be detected on it. The action here described, by means of which an electrified body can excite the electric condition in a neighboring body, which is insulated from it, is called *electric induction*.

The attraction between a charged body and any other body brought near it is due to the attraction between the original charge of the first body and the neighboring charge of the opposite kind induced in the second body. When the second body is insulated and contact occurs between the bodies, this charge of the opposite kind is destroyed by union with some of the original charge and both bodies become charged similarly. In this condition they repel each other.

416. Free and Bound Charges. — If a conductor, when in the presence of a charged body and charged by induction with both kinds of electricity, is touched by the hand, or is connected with the earth by any conductor (Fig. 243), the charge which it has on the end farthest from the electrified body, and which is of the same sort as that of the electrified body, disappears. The other charge, on the end nearest the electrified body, remains unchanged in character, though its distribution on the conductor will be altered. These two charges, therefore, seem to differ, in that one of them is free to leave the conductor by passing through any conductor presented to it, while the other is bound or retained in the conductor. They are called the *free* and the *bound* charge respectively. The experiments here detailed show that the bound and the free charges are equal in magnitude, and that the bound charge is the one which is opposite in kind to the original charge.

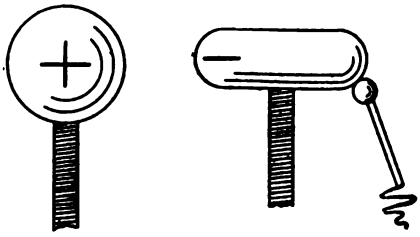


Fig. 243.

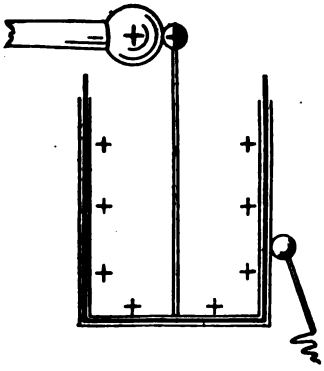


Fig. 244.

417. The Leyden Jar. — To apply these facts to the explanation of the behavior of the Leyden jar, we notice that, when the jar is being charged (Fig. 244), the inner coating receives a charge directly from the machine, and acts as the original electrified body. By the action of the charge which it

first receives, the outer coating is charged by induction, and its induced free charge passes off to the earth. Its bound charge acts by induction on the inner coating so as to make a free charge on that coating of the opposite sort to that which is supplied by the machine. The machine therefore can supply an additional charge to neutralize this, and the charge on the inner coating is increased by the amount of the bound charge which was developed in it. This increased charge on the inner coating again acts by induction, to increase the bound charge on the outer coating, and this process goes on until the jar discharges itself or until a limit is reached, which is determined by the condition of the electric machine that is supplying the charge. In the final condition of the jar when charged, the conductor of the machine and the knob of the jar are in similar electric conditions, so that no farther charge can pass between them, and two opposite charges, of practically equal magnitude, confront each other on the surfaces of the two coatings which are next to the glass. When conducting connection is made between the outer and the inner coating, these charges recombine through the conductor, and the jar is discharged.

418. The Gold-Leaf Electroscope. — The gold-leaf electroscope is an instrument used for detecting the presence of electric charges and for determining their nature. It consists of a pair of parallel strips of gold foil suspended from the end of a short metallic rod, which is carried in the stopper of a small jar (Fig. 245). To prepare the instrument for use, a negatively charged body is brought over the knob in which the rod terminates, and induced charges are set up in the rod and the dependent strips of foil. The strips are charged negatively, and as they repel each other, they diverge from each other. The rod is then touched by the hand, so that the free charge, which in this case is the negative charge, is withdrawn, and the repulsion ceasing, the strips no longer diverge from each other. On withdrawing the hand, and then withdrawing the external negative charge, the bound positive charge is distributed over the conductor, and the gold leaves again diverge, being positively charged. The instrument is then ready for use. A positively charged body, brought near the knob, will induce additional positive charge on the gold leaves, and they will diverge more widely. A

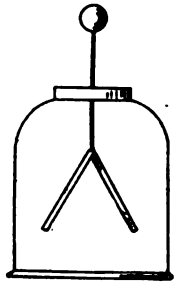


Fig. 245.

negatively charged body will induce a negative charge on the gold leaves, and their divergence will diminish.

419. The Electrophorus. — The laws of electric induction are also illustrated in an instrument invented by Volta (1775), known as the *electrophorus*. It consists (Fig. 246) of a sheet of sulphur, rubber, or some similar substance, which can be electrified by friction, called the *plate*, and a metallic disk furnished with an insulating handle, called the *carrier*. The plate is electrified by friction, generally negatively, and the carrier placed upon it. In this position the carrier is usually supported on slight prominences in the plate, so that it is in contact with the plate at very few points, and does not receive a charge from it by conduction. It is charged by induction, the charge on its lower surface being positive, opposite to that of the plate. The negative charge on its upper surface is free, and is removed by touching the carrier with the hand. When the carrier is lifted from the plate, the bound positive charge, being no longer constrained by the presence of the negative charge of the plate, distributes itself over the carrier. This charge may then be communicated to any conductor which it is desired to charge, and the carrier may be charged again in a similar manner, as often as we please, the original charge on the plate being retained by it and remaining unaltered by the operation.

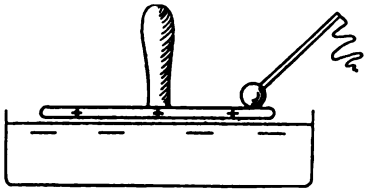


Fig. 246.

420. Distribution of Electricity. — On the hypothesis that the electric charge is some sort of fluid whose parts repel one another, it is easy to see that the distribution of electricity in an insulated conductor will depend upon the shape of the conductor. It was shown first by Franklin (1749), and long afterwards more conclusively by Faraday, that whatever be the shape of the conductor, the charge in it resides entirely on its surface. By testing different bodies with the proof-plane, it was also shown that the density of the charge, that is, the amount on unit of surface, is greatest at the parts of greatest curvature. When the law of electric force was discovered, the distribution on bodies of particular forms was calculated. The experimental examination of the distribution on such bodies verified the results of these calculations.

The density of the charge is especially great at any part of the conductor which ends in a sharp point. At such points the pressure of

the dense charge against the surrounding insulator is especially great, and the charge therefore passes from them with great facility. This property of points, of discharging bodies, and also of receiving charges, is made use of in all sorts of electric apparatus.

421. Early Theories of Electricity. — After Dufay had discovered the opposite effects produced by the charges on glass and on resin, he developed a theory of electricity to explain them. He assumed that all matter contains two electric fluids, which he called vitreous and resinous electricity; that in any uncharged body these fluids exist in equal amount; that they are separated from each other by friction, so that one of the two bodies which are rubbed together retains a surplus of the one fluid, while the other body has a surplus of the other fluid; that the parts of either of these fluids repel other parts of the same fluid and attract parts of the other fluid. This theory, which is commonly known as the two-fluid theory of electricity, was not at first accepted, but it gradually won its way into prominence, and is the one which has been used in most of the development of the subject.

Franklin adopted the view that there exists in the universe a single fluid, called electricity, and that a body when charged positively possesses a surplus of this fluid, while a body charged negatively possesses less than the normal amount. This theory was developed by Æpinus, and for some time was generally held. It fell into disuse after a while, in consequence chiefly of the necessity which Æpinus found, in order to explain the electric attractions and repulsions, as well as the passivity of unelectrified bodies, of supposing that matter repelled matter according to the same law as that by which electricity repelled electricity. Without the assumption of a mutual repulsion between the particles of matter, as well as of an attraction between matter and electricity, the actual forces observed could not be explained.

The modern theories of electricity unite in assuming that electricity, whatever it may be, is not a continuous fluid, but exists in separate portions, or units, which may be called electric atoms or electrons.

422. Law of Electric Force. — For any further development of the subject, it becomes important to know the law of the force exerted by one electric charge on another.

This law was first investigated by Cavendish (1773). His method of investigation depended on a theorem first proved by Newton (Examples, XI, 1) to the effect that the strength of field at any point within a uniform spherical shell, com-

posed of any substance exerting force which varies with the distance according to the law of inverse squares, will be zero. By the most careful experiments which he could make with the crude apparatus at his disposal, Cavendish could not detect any evidence of any electric force inside a sphere charged with electricity. He therefore concluded that the elements of electricity act on one another with forces which vary inversely with the square of the distance. The determination of this law, by this method, involves the assumption that the force exerted between two quantities of electricity is proportional to their product. Cavendish never published his proof of this law, and it remained unknown until his papers were edited by Maxwell. The general method used by Cavendish has been employed with the most sensitive apparatus, and the law of electric force has been determined with great exactness.

The law of electric force became known from the work of Coulomb (1785). In his investigations, Coulomb employed the torsion balance already described in § 383. By means of this instrument he measured the repulsions between two similarly charged spheres at different distances from each other, and found that the force between them varied inversely with the square of the distance between their centers. It had been proved by Newton (Examples, XI, 2) that the force at any point outside a spherical shell made of a substance which exerts force according to the law of inverse squares will be inversely as the square of the distance of the outside point from the center of the sphere. The measurements of Coulomb were therefore consistent with the hypothesis that each element of the electric charge repels every other element with a force which varies inversely as the square of the distance. A similar law was proved for the attraction between two spheres oppositely charged.

By observing the force between two electrified spheres at a definite distance, and comparing it with the force exerted at the same distance between the spheres when the charge on one of them had been halved by touching the sphere to another of the same size, Coulomb showed that the force is proportional to the magnitudes of the charges.

The law of electric force may therefore be stated by saying that the force between two small electric charges is proportional to the product of the charges and inversely proportional to the square of the distance between them. Owing to the fact that the force between two charges depends on the medium which surrounds them, we assume in this definition that the charges are in a vacuum. The force between the charges in air is only slightly smaller than that in vacuum.

423. Quantity of Electricity. — In measuring quantities of electricity we find ourselves in the same position as when we measure quantities of magnetism. We do not perceive electricity directly, or by any other effects than the forces which it exerts. We are therefore compelled to measure it by means of those forces. If we represent two electric charges by e and e', the distance between them by r, and the factor of proportion by k, the equation which expresses the force between the charges is

$$F = k\frac{ee'}{r^2}. \tag{129}$$

We can measure the force and the distance, but until some convention is made, we cannot measure the charges. It is conceivable that a charge might be selected arbitrarily as a standard, with which other charges might be measured by the examination of the forces which they exert on the standard charge. But the impossibility of preserving such a charge unaltered, and of using it in the study of other charges without error, makes it necessary for us to determine the unit charge in some other way. If we were to use this method, the factor k would have a definite value, which might be determined by experiment as the constant of electric force. But so long as the unit charge is left undetermined, the value of k is not fixed. By assuming an arbitrary value for it, we fix the product of the two charges which act on each other. To make this value as simple as possible, we set $k = 1$. When this choice is made, the value of the product ee' is determined from measured values of the force and the distance. If the charges are equal, the value of either charge is determined from the same data. With this convention the force between two charges is represented by the formula

$$F = \frac{ee'}{r^2}. \tag{130}$$

424. Unit Electric Charge. — In particular, we may define the *unit electric charge* as that charge which will repel an equal and similar charge at unit distance with unit force. In the c.g.s. electrostatic system the unit electric charge is one which will repel an equal and similar charge at the distance of one centimeter with the force of one dyne.

In the practical system of units used in most scientific and technical work, the unit electric charge is called the *coulomb*. It is properly defined by reference to electric phenomena with which we are not ready to deal at this point; but it may be taken, without

definition, as approximately equal to 3×10^9 electrostatic units of charge.

425. Electric Fields. — In order to study the electric forces exerted by any system of electrified bodies it is convenient to suppose that we can make use of a unit charge, which is perfectly insulated, and the force upon which we can always measure. This charge may be called a *test unit.* When such a test unit is brought near an electrified body, it is generally acted on by an electric force. The region in which such a force can be detected is called an *electric field.*

426. Electric Intensity or Strength of Field. — If a test unit is placed at any point in an electric field, the magnitude of the force which acts on it is called the *electric intensity* or the *strength of field,* or sometimes simply the *electric force* at that point. We may designate it by the symbol R. The force exerted at any point on a charge e is then given by eR.

The field of unit intensity is one in which the test unit is acted upon by unit force. In the c.g.s. electrostatic system it is the field in which the test unit is acted upon by the force of one dyne.

The intensity, from its definition, is plainly a vector, and may be resolved into components, or two intensities may be compounded to find a resultant, by the parallelogram law.

427. Lines of Electric Force. — If a test unit is placed at any point in the electric field, and is then allowed to follow the direction given to it by the electric force, it will trace out a continuous curve. This curve is called a *line of electric force.*

Defined geometrically, a line of electric force is a line so drawn in a field of electric force that the tangent to it at any point indicates the direction of the force at that point.

The positive direction of a line of force is the direction in which a positive charge will tend to move along it.

428. Electrification by Induction. — When an insulated conductor is placed in an electric field, it becomes *electrified by induction.* We may describe the distribution of the charge on it, in a general way, by saying that both the positive and the negative charges move in the conductor under the action of the electric field as if they were free. The positive charge is therefore displaced in the positive direction of the lines of force, the negative charge, in the negative direction. This displacement continues until the forces set up by the displaced charges, combined with the electric force of the field, neutralize one another at every point within the conductor. When

this condition is reached, the displacement of the charges ceases and the system is in electric equilibrium.

429. Electric Potential. — In our study of magnetism we have already defined a function called the magnetic potential, and used it to describe the characteristics of various magnetic fields. This function was first introduced by Laplace for the purpose of using it in the description of fields of gravitational force. Green applied it to the study of problems in electricity, and gave it the name of potential function.

The theorems in which the properties of the potential are stated, in so far as they depend only on the law of the force which gives rise to the field, are the same for magnetic, gravitational, and electric potential; but owing to the peculiar mobility of electricity on and through conductors, the description of the electric state of conductors, and of the electric field around conductors, can be made in the case of electricity, by the aid of this function, with special simplicity.

If the electric field is due to a single positive electric charge e on a conductor which is so small that the charge may be considered as located at a point, and if a test unit charge is moved in the field from a point A at the distance r from the charge to a point X at the distance x from the charge, the work done on the test unit by the force in the field will be expressed by $e\,(1/r - 1/x)$. The proof of this is the same as that used already in Examples, XI, 6, or in § 391. If the test unit moves off to an infinite distance, the work done is expressed by e/r. This quantity measures the *potential* V_A at the point A. The work done on the test unit as it passes from the point A to the point X is $V_A - V_X$, the difference of potential between A and X. The difference of potential may be shown, as in Examples, XI, 7, to be independent of the path pursued by the test unit, and to depend only on the distances of the initial and terminal points of the motion from the central charge.

If the field is due to many electric charges the difference of potential between two points A and X will be expressed by

$$V_A - V_X = e_1\left(\frac{1}{r_1} - \frac{1}{x_1}\right) + e_2\left(\frac{1}{r_2} - \frac{1}{x_2}\right) + \cdots e_n\left(\frac{1}{r_n} - \frac{1}{x_n}\right), \qquad (131)$$

or by $\Sigma e\left(\frac{1}{r} - \frac{1}{x}\right)$. The potential V_A is expressed by

$$V_A = \sum \frac{e}{r}. \qquad (132)$$

The difference of potential $V_A - V_X$ will be independent of the path pursued by the test unit, and will depend only upon the positions of the initial and terminal points of the motion.

From the derivation of the quantities expressing the difference of potential and the potential we can give the following definitions:

The difference of electric potential between two points in a field of electric force is measured by the work done upon a test unit by the forces of the field as the unit moves from one of the points to the other.

Also, the potential at a point in a field of electric force is measured by the work done upon a test unit by the forces of the field as the unit moves from that point to an infinitely distant point.

430. Equipotential Surfaces.— If we set $V_A = \sum \frac{e}{r} = \text{constant}$, we may satisfy this equation by continuous variations of the values of r, and may thus determine a surface on which the potential has the same value at every point. Such a surface is called an *equipotential surface*. Since no work will be done on the test unit if it is moved about in an equipotential surface, there can be no component of electric force in the surface, so that the equipotential surface is everywhere perpendicular to the direction of the electric force. Therefore, if we have the electric field mapped out by its lines of force, we may describe an equipotential surface by starting at any point and drawing the surface through that point which is everywhere perpendicular to the lines of force; and if further we choose the points through which such surfaces are drawn at such distances apart along a line of force that one unit of work is done on the test unit as it moves from one such point to the next one, we may map out the whole field in a systematic way by equipotential surfaces, characterized by the property that the potential difference between any two consecutive ones shall be unity.

431. Relation between Force and Potential. — If the test unit is moved through the very short distance s from the point A to the point B, the work done on it is expressed by $V_A - V_B$. Within the short distance covered by the motion we may suppose the component of force in the direction of the motion to be constant. Calling this component F, we have again for the work done on the test unit $Fs = V_A - V_B$; and hence

$$F = \frac{V_A - V_B}{s} = -\frac{V_B - V_A}{s}. \qquad (133)$$

We define the limit of the ratio $\frac{V_B - V_A}{s}$, when s is drawn from

A to B, as the rate of change of the potential with respect to space. Using this definition, we conclude that the force at a point in any direction equals the rate of change of potential in that direction taken with the opposite sign.

If the line s is drawn in an equipotential surface, $F = 0$, or the force has no component in the surface. The force is perpendicular to the equipotential surface. If we know by observation that $F = 0$ in every direction, we know that $V_A = V_B$, or that the potential is constant in the region about the point.

The force F does positive work if V_A is greater than V_B. The test unit, or any positive electric charge placed in the field, will therefore tend to move from a point of higher to a point of lower potential. A negative charge will tend to move in the opposite direction.

432. The Potential of a Conductor. — The reason for the peculiar applicability of the potential in the discussion of problems in electricity is the fact that electricity moves freely in a conductor under the influence of electric force, so that the final condition of an insulated charged conductor is one in which there is no electric force at any point within it, and the electric force at its surface is everywhere perpendicular to the surface. From the theorem just proved (§ 431), this means that the surface of the conductor is an equipotential surface, and that all points within it are at the same potential as the surface. The potential which all points of the conductor have in common is called the *potential of the conductor*.

We may illustrate the application of the idea of potential by describing the use of the electrophorus. When the plate of the electrophorus is charged by friction, its potential, and that of the region around it, becomes negative. The numerical value of this negative potential is highest at the plate, and diminishes from the plate to zero. When the carrier is placed on the plate, it is brought into this region of negative potential, and if it were not a conductor, the potential on its lower surface would be negatively higher than the potential on its upper surface. There would thus exist in it a difference of potential, and the lines of force corresponding to this difference would be directed from its upper surface to its lower surface. Such a force, however, cannot exist in a conductor, and a positive charge is developed on the lower surface, and a corresponding negative charge on the upper surface, until the forces to which they give rise neutralize the force which has been described, or until the potential within the carrier is everywhere the same. When the carrier is touched and so joined to earth, its potential is raised to the potential of the earth. This involves the loss of the negative charge of the carrier. The potential of the carrier is then due to the superposition of the opposite potentials due to the charge on the plate and the positive charge on the carrier. When the carrier is lifted, work is done on it against the attraction between these charges, and the work thus done raises the electric energy of the carrier, and gives it positive potential.

433. Zero Potential. — In order to measure the potential of a conductor according to the definition of it which has been given, it would be necessary to carry a test unit from that conductor to infinity. By this operation we measure the difference of potential between the conductor and an infinitely distant point. Of course this operation is impossible, and it is therefore necessary, for practical purposes, to adopt the potential of some other point as the standard potential, with which other potentials shall be compared. The standard potential chosen is the potential of the earth, which, it may be shown (Examples, XXIV, 4), remains appreciably constant at all times, whatever be the electric operations which take place during our experiments. This standard potential we take as the *zero of potential.* With this convention the potential of a conductor is the work that is done on a test unit as it moves from the conductor to the earth.

434. Unit Potential. — The unit of potential is the potential of a conductor so charged and so situated with respect to other conductors that one unit of work is done on a test unit charge as it moves from the conductor to the earth. In the c.g.s. electrostatic system it is the potential of a conductor so charged and so situated with respect to other conductors that one erg of work is done on a c.g.s. electrostatic unit of electricity as it moves from the conductor to the earth. This unit is too small to be convenient for general use.

In the practical system of electric units used in most scientific and technical work the unit of potential is called the *volt.* The volt is defined as the potential of a conductor so charged and so situated with respect to other conductors that one joule of work (10^7 ergs) is done on a coulomb of electricity (§ 424), as it moves from the conductor to the earth. Differences of potential are often measured in volts by reference to the difference of potential between the two terminals of a standard voltaic cell (§ 503).

435. Electric Machines. — The first electric machine, constructed by von Guericke, has already been described. It was very soon improved upon by using a glass plate or cylinder, instead of the sulphur ball, and by setting against it a pad of flannel or leather, by the friction of which the glass is electrified. An insulated conductor, called the prime conductor, furnished with a comb or row of points, pointing toward the glass, is used to collect the charge developed as the machine is turned. Machines of this sort are called frictional machines. They are now very little used.

The germ of all modern electric machines is found in the electro-

phorus of Volta. It is plain that if the operation of charging and discharging the carrier, which we have described as carried on by the hand, is executed rapidly by some mechanism, a rapid succession of charges can be obtained. It is not necessary to go into the details of these induction machines.

436. Electroscopes and Electrometers. — Any apparatus which will indicate the presence of an electric charge, and which will enable us to determine its character, is called an *electroscope*. Any light body, such as a straw suspended by a silk thread, may serve for that purpose. The gold-leaf electroscope, which is a very common form of the instrument, has been described in § 418.

An *electrometer* is an instrument by which the difference of potential between two bodies can be measured, or compared with some other difference of potential. The attracted-disk electrometer, first used by Snow Harris and developed by Lord Kelvin, furnishes a measure in absolute units of difference of potential. It consists essentially of a large horizontal disk, above which a smaller disk is set at a known distance. This smaller disk is supported by a balance or a spring, by means of which the force upon it may be measured. By measuring the force exerted by the charges on the disks, when they are brought to different potentials, the difference of potential between them may be measured in absolute units. (Examples, XXIV, 3.)

EXAMPLES, XXIII

1. *To calculate the potential of a spherical conductor when no other conductors are near it (a freely electrified sphere).*

The radius of the sphere is r, the charge on it e. The potential of any part of it is the same as that at its center. The charge is uniformly distributed over the surface and each element of it will set up a potential at the center equal to its amount divided by the radius. The sum of all these potentials, or $V = e/r$, is the potential of the sphere.

2. *To show that the potential at a point outside a uniformly charged sphere is equal to the charge divided by the distance from the point to the center of the sphere.*

As will be shown in Example 4, the force at the point is equal to e/R^2, where e represents the charge and R the given distance. The same demonstration as that used in Examples, XI, 6, will show that the potential is then given by e/R.

3. *To prove that the electric force at a point within a charged sphere is equal to zero.*

See Examples, XI, 1.

4. *To prove that the electric force at a point outside a charged sphere is inversely proportional to the square of the distance of the point from the center of the sphere.*

See Examples, XI, 2.

5. *To prove that the total electric flux taken over a closed surface which does not enclose any electricity is equal to zero.*

We define the electric flux over a small area s, at each point of which the electric force has the same value R, as equal to $Rs\cos\alpha$, where α is the angle between the normal to the surface and the direction of the force. The conventions used in the demonstration are the same as those made in the analogous case of magnetism.

We use the same proof as that of Examples, XXI, 7, for Gauss's theorem.

6. *To prove that the total electric flux over a closed surface that encloses electric charges is equal to $4\pi\Sigma e$, where Σe is the sum of all the charges within the surface.*

The proof is the same as that used in § 396.

7. *To show by Gauss's theorem that the force within a uniformly charged sphere is zero.*

Within the charged sphere, describe a sphere concentric with it. This sphere contains no electricity, and therefore the total electric flux over it is zero. But from symmetry, the electric force at every point of it should be equal and similarly directed inward or outward from the center, and thus, since the flux can contain only terms of one sign, and since the surface of the sphere is not equal to zero, the force at every point of its surface must vanish.

8. *To show by Gauss's theorem that the force at a point outside a charged sphere equals the charge on the sphere divided by the square of the distance from the point to the center of the sphere.*

Through a point at the distance R from the center of the charged sphere, describe a sphere concentric with the charged sphere. Then the force F at every point of it, by symmetry, will be the same and similarly directed inward or outward from the center C, and the total electric flux over its surface will be $4\pi R^2F = 4\pi e$, whence $F = e/R^2$.

9. *To find the electric intensity at a point infinitely near a charged body.*

The surface density of electric distribution is defined as the ratio of the charge on an area to the area. It is commonly represented by the symbol σ. If the charge is e and the area a, we have $\sigma = e/a$, if the distribution is uniform; or $\sigma =$ the limit of the ratio e/a, as the area a diminishes indefinitely, if the distribution is not uniform.

The surface of the charged body is an equipotential surface, so that the lines of force are perpendicular to it. We describe a closed surface enclosing a small area a of the surface on which the surface density of distribution is σ, by drawing (Fig. 247) two surfaces A and B parallel to the chosen area, and very near it, one inside and one outside the body, and a tubular surface perpendicular to the surfaces A and B and terminating on them as cross sections. This closed surface encloses σa units of electricity, and by Gauss's theorem the total electric flux over it is $4\pi\sigma a$. There is no force at any point of the surface within the body; the force at every point of the tubular surface outside the body lies in the surface and contributes nothing to the total flux; the force at any point P in the surface B outside the body is the same if the area a is small enough, and so the total flux is $Fa = 4\pi\sigma a$, whence $F = 4\pi\sigma$.

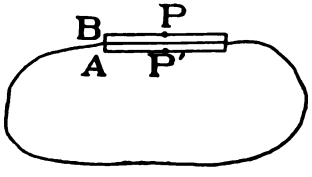


Fig. 247.

10. *To find the force at a point near the area a of Example* 9, *due to the charge on it alone.*

The force at the point P outside the charged body is $4\pi\sigma$; that at a point P' (Fig. 247) inside the body is 0, so that the force changes by $4\pi\sigma$ as we pass through the area a. This change is due to the reversed direction of the force due to the charge on area a, and this force is therefore equal to $2\pi\sigma$.

11. *To find the force exerted by a uniformly charged plane sheet on a circular plane sheet placed near it, and parallel with it, when the charge on the second sheet is of the same density but of opposite sign.*

Let the area of the circular sheet be a. The charge on it is σa. Each unit of charge on it is attracted by the other sheet with a force $F = 2\pi\sigma$ (Example 10); so that the total force exerted on the circular sheet is $2\pi\sigma^2 a$.

12. *To find the force exerted on a unit of area of a charged conductor by the charge on the conductor (the electric pressure).*

Consider an area a taken in the surface of a charged body and so small that it may be treated as a plane area. In Example 10 we proved that the force at a point near that area due to the charge on it is equal to $2\pi\sigma$. Since the force at the same point due to the charge on the whole body is $4\pi\sigma$, the force at that point due to all the charge except that on the area is $2\pi\sigma$. This force acts on each unit of charge on the area, and therefore the force F on the area is $F = 2\pi\sigma^2 a$, and the force on unit area is $F/a = 2\pi\sigma^2$. This force acts outward from the surface of the body, whether the charge is positive or negative.

13. *To calculate the energy expended in charging a conductor.*

The proof is similar to that given in Examples, XXI, 8. The operation of charging may be carried out by bringing up small charges, each equal to $1/n$th the final charge E, successively to the conductor from infinity. The conductor is originally at zero potential. When it receives the first elementary charge on which no work is done to bring it to the conductor, its potential rises to $1/n$th the final potential V. The work done in bringing up the second charge is therefore $(E/n)\ (V/n)$; and the potential rises to $2V/n$. The work done in bringing up the third charge is $(E/n)\ (2V/n)$, and so on. The work done in charging the conductor is therefore

$$EV\frac{(1+2+\cdots(n-1))}{n^2} = \frac{EV}{2}\frac{(n-1)n}{n^2} = \frac{EV}{2}\frac{(n-1)}{n} = \frac{EV}{2}$$

if n is a very great number.

The quantity $EV/2$ also represents the energy of the charge on the conductor.

437. Equality of the Two Kinds of Electricity. — It was assumed by Dufay and Franklin, and accepted by all other students of the subject, that the two electric conditions, called vitreous and resinous, or positive and negative, exist in every natural body in equal quantities. Our attention has already been called to this assumption in our study of induction, in which we saw that the two charges which are separated by induction will exactly neutralize each other when the conductor is removed from the electric field.

It may also be shown by experiment that when electricity is produced by friction, it appears in equal quantities of the two kinds.

For example, if a rod of glass is fitted with a cap of silk or flannel, held in an insulating handle, and is turned in the cap until its end is electrified, it may be shown that the cap is also electrified oppositely; and the equality of the two charges is shown by the fact that when the cap is on the glass rod, no electric effects can be detected. This experiment succeeds best if the glass and silk are rubbed together inside a closed metallic vessel insulated from the earth. As we shall see in the next paragraph, the exterior of this vessel ought to appear electrified unless the two charges developed are exactly equal. That two opposite charges are developed is shown by removing either the glass or the silk from the vessel, the exterior of which then becomes either negatively or positively electrified. But when the glass and silk are in the vessel together, no external electrification can be detected. Indeed, any electric machine, or the most elaborate methods for developing electricity, may be worked inside the vessel without electrifying its exterior at all.

The final demonstration of the equality between the positive and negative charges, and of the invariable relation between them, was given by Faraday (1843). The experiment by which Faraday did this is commonly called the *ice-pail experiment*, because the vessels which he used in it were ice pails. In its simplest form, the apparatus consists of a metallic vessel, set on an insulating stand, and of an insulated conductor, which can be introduced into the interior of the vessel. The cover of the vessel is put on after this conductor is introduced. Some sensitive apparatus, such as a gold-leaf electroscope, is provided, by which the electrification of the exterior of the vessel can be examined. The vessel is first discharged by joining it to the earth for a moment, and the conductor, charged positively,

a Fig. 248. b

is introduced within it. The vessel is then found to be charged positively (Fig. 248, a). If it is then joined to earth for a moment, its positive charge disappears. If the charged conductor is now removed from the interior, the vessel exhibits a negative charge. (Fig. 248, b).

Judging from what we have already learned about the distribution of induced charges, we may conclude that the effect of introducing the charged conductor within the vessel is to develop a negative charge on its inner surface and a positive charge on its outer surface.

These charges are equal in magnitude, as is proved by their neutralizing each other if the charged conductor is removed before the positive charge of the vessel has been removed by joining it to earth.

Now to determine the relative magnitudes of the original charge on the conductor and the opposite charge which it produces by induction, we introduce the charged conductor as before, and remove the induced positive charge. We then touch the charged conductor to the interior of the vessel, or join it to the interior by a conducting wire. No signs of electrification appear on the exterior (Fig. 249), and now when the conductor is removed, no charges can be detected either on it or on the vessel. Faraday concluded from this result that the charge on a body induces a charge of the opposite sort, and of equal magnitude to itself, on the conductors which completely surround it.

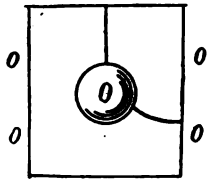


Fig. 249.

Before the significance of this experiment was appreciated, it was usual to speak of charged bodies which are at some distance from other conductors as *freely electrified* bodies, and to ignore the opposite charges which they produce in neighboring bodies by induction. It was only when the charged body is very near other conductors, as in the case of the Leyden jar, that the presence of the induced charge was thought of as having any special significance. Faraday's experiment showed, however, that there is really no such thing as a freely electrified body, but that the electrification of one body involves the electrification by induction of the conductors which surround it. This general truth suggests the hypothesis that the two equal charges which confront each other across the nonconducting medium which separates them, are the two ends of a condition which exists in the medium, and that the electrification of a body involves setting up this condition in the medium. We shall find additional support for this hypothesis when we study the way in which different media affect the charge which a body receives, when the conditions are otherwise the same.

438. Capacity of Conductors. — We define the *capacity* of a conductor as the quantity of electricity which is required to raise the potential of the conductor from zero to unity, when the potential of all surrounding conductors is maintained at zero. Since the potential at any point, if the charges which give rise to the field are simply changed in magnitude in the same proportion, without being changed in position, will change in proportion to the charges, the

capacity will be also equal to the ratio of the charge on the conductor to its potential. Representing the capacity by C and the charge by e, we have

$$e = CV. \tag{134}$$

The capacity of a conductor depends, at least in part, on its shape and size, and on the distance between it and surrounding conductors. If we examine the explanation which has been given in § 417 of the charging of a Leyden jar, it will be seen that the mutual inductive action between the two coatings of the jar will be greater when they are nearer together, and that consequently a larger charge will be received by the inner coating from the same source, when the coatings are near together, than when they are farther apart.

This general conclusion may be formally demonstrated in the case of the spherical condenser. The spherical condenser (Fig. 250) is a conducting sphere insulated from a concentric spherical conducting shell. The outer shell is joined to earth, and through a small hole in it contact is made by means of a wire between the interior sphere and the source of charge. The potential of the interior sphere is everywhere the same, and we can calculate it by calculating the potential at its center. Now all parts of the charge on the interior sphere of radius r are at the same distance from the center of the sphere, and the potential at that center due to the charge on the sphere is equal to e/r. The potential at the same point, due to the equal and opposite charge on the inner surface of the exterior shell of radius r' is $-e/r'$. The actual potential at the center of the sphere, and therefore of the sphere itself, is the sum of these potentials, so that we may write $V = e\left(\frac{r'-r}{rr'}\right)$. The ratio of the charge to the potential, or $\frac{rr'}{r'-r}$, is the capacity of the interior sphere as we have defined it. As the formula shows, it is equal to the product of the two radii divided by their difference, and therefore increases very rapidly as the difference of the radii is diminished.

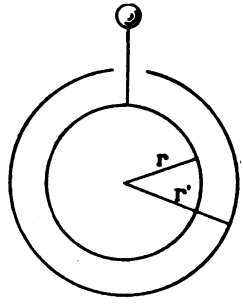


Fig. 250.

A system of conductors like the spheres just described, or like the Leyden jar, of which the capacity is very great, is called a *condenser*. The conductor which is charged in a condenser differs from other conductors merely in being so situated, with respect to the conductors which surround it, that a relatively large charge is required to raise it to unit potential.

It is obvious that the conclusion which has been proved to hold for the spherical condenser, that its capacity depends upon the distance between the charged body and the conductors which surround it, will hold for conductors of any shape.

439. The Dielectric. — An isolated experiment tried by Franklin led him to believe that the nonconducting medium between the two coatings of a Leyden jar plays an important and perhaps the principal part in the charging of the jar. We now try this experiment with what is known as a dissected jar, that is, a glass vessel or cup which is set inside a metallic cup serving as an outer coating, and receives another metallic cup, furnished with a conducting rod, serving as the inner coating. Such a jar may be charged and discharged in the ordinary way. If the inner coating is removed while the jar is charged, it will be found to have only a very small charge on it. It appears, therefore, that the charge of the jar is not carried about on the inner coating. If this coating is replaced, the full charge may be obtained from the jar. The experiment is more striking if we remove both the coatings and replace them by others of similar shape. The jar is still found to be charged, and the conclusion is irresistible that the charge has resided somewhere in or on the glass. It is natural to make the hypothesis that the process of electrification consists in setting up in the glass some peculiar condition, which terminates at the two conducting coatings; though as the surfaces of the glass are often at least partial conductors and as the charge may reside on them, this hypothesis is by no means demonstrated by the experiment.

When a nonconducting medium is used to separate a charged conductor from the other conductors which surround it, and when we are studying the conditions which exist in that medium, or the influence of the medium upon the charge, we call the medium the *dielectric*. The first study of the properties of dielectrics was made by Cavendish (1771–1781), but his work was not published at that time, and we owe our knowledge of those properties to Faraday (1837). Faraday's investigation was carried out by the aid of two precisely similar spherical condensers (Fig. 251). In one of these the space between the two surfaces was always filled with

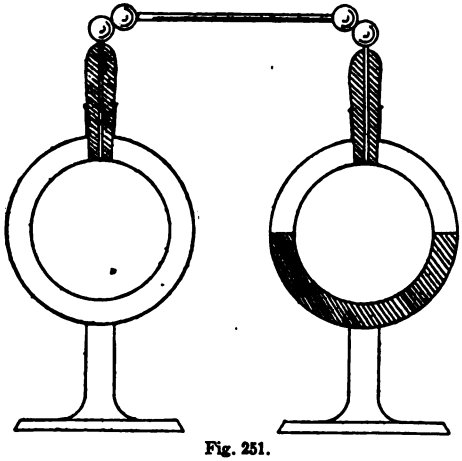


Fig. 251.

air, in the other it was first filled with air, and afterwards the lower half was filled with the different dielectrics which he studied. The experiment consisted in charging the inner spheres of both these condensers to the same potential, by joining them both at the same time to the same source, and in discharging them in turn in such a way that the quantities discharged could be compared. As was to be expected from the similarity of the condensers, the quantity discharged was the same from each when both contained air as the dielectric. When sulphur or shellac was used in one of the condensers as the dielectric, the quantity obtained from it was considerably greater than the quantity obtained from the other. This experiment proved that the capacity of a condenser depends on the nature of the dielectric which separates its conducting parts. Faraday used the term *specific inductive capacity* to represent the effect of a particular dielectric, as determined by the ratio of the capacity of a condenser in which it is used to the capacity of a similarly shaped condenser in which air is used. This ratio, which is a characteristic constant for each particular dielectric, is now generally called the *dielectric constant.*

If we represent the dielectric constant of a medium which separates a conductor from the surrounding conductors by K, the charge e on it will be connected with the capacity C which it will have if surrounded by air, and with the potential V, by the formula

$$e = KCV. \tag{135}$$

This direct and very complete evidence of the participation of the dielectric in the process of charging a body confirmed the hypothesis which has already been referred to, that the production of the charge involves setting up in the dielectric a peculiar condition, which terminates at the charged conductors. This hypothesis has been worked out analytically by Maxwell, and has been shown to give a complete account of the relations of charged bodies.

440. Tubes of Electric Force. — The many points of analogy between the electric field and the magnetic field make it at once evident that we can apply to the electric field the mode of description which serves so well for the magnetic field, and consider the charges on conductors as due to a condition of the nonconducting medium, or dielectric, between them. This view was held by Faraday, who described the electric field by means of a proper distribution in it of physical lines of electric force, and accounted for the forces between electrified bodies by stresses in these lines of force.

If a small area is taken on an equipotential surface in a field of electric force, and if the geometrical lines of force which pass through its boundary are traced throughout the field, it will be found that, owing to the continuity of the force in the field, those lines which lie contiguous to one another as they pass through the boundary will lie contiguous to one another throughout the field, so that they will mark out a tubular surface. The region enclosed by this surface is called a *tube of electric force*. When the dielectric is homogeneous, these tubes begin and end on conductors, as is proved by Faraday's ice-pail experiment (§ 437).

441. Flux of Electric Force. — The flux of electric force through a surface so small that the force at all parts of it has the same value, is defined as the product of the normal component of the force and the area of the surface. The flux of electric force through any surface is then the algebraic sum of the fluxes through the elements of area making up that surface.

The theorems concerning electric flux are similar to those concerning magnetic flux. In particular, the flux of electric force is constant throughout a tube of force. To show this we consider a portion of a tube of force bounded at its ends by the areas cut out by the tube from two equipotential surfaces, these areas being taken so small that the force at all points of one of them has the same value. Let R_1 and R_2 represent the forces at the two end surfaces, s_1 and s_2, the areas of those surfaces. The contributions to the total flux by the end surfaces are $R_1 s_1$ for one, $R_2 s_2$ for the other, and they are of opposite sign. There is no flux through the wall of the tube, since the lines of force lie in it. By Gauss's theorem (Examples, XXI, 7), the total flux, $R_1 s_1 - R_2 s_2 = 0$, or

$$R_1 s_1 = R_2 s_2. \tag{136}$$

The electric flux is therefore the same throughout the whole length of a tube of force.

Since this proposition is true for a narrow tube, in which the force at all points of a cross section is the same, it is also true for all the tubes of force which leave a charged body. If a closed surface is described in the field so as to enclose a charged body, the flux through the surface is the same, wherever it is placed and whatever its shape may be.

442. Description of Field by Tubes of Force. — If a closed equipotential surface is described in a field of electric force, and if it is divided into small areas, so taken that the flux of force through each

one of them is equal to 1, the tubes determined by these areas are *unit tubes*. The whole field is filled with these tubes. If s is the area of the section of a unit tube by an equipotential surface at any point, $1/s$ is equal to N, the number of such unit tubes which would pass through unit area of the equipotential surface at that point, and since the force R at the point is such that $Rs = 1$, we have $R = N$. The strength of field at any point is therefore equal to the number of unit tubes of force which pass through unit area of the equipotential surface at that point. The whole field may therefore be mapped out by means of its unit tubes of force.

443. Relation of Tubes of Force to Charge. — Let us suppose an equipotential surface drawn enclosing the electric charge e, and let the unit tubes be drawn which pass through that surface. By Gauss's theorem (Examples, XXI, 7), the total flux through the closed surface is $4\pi e$, and since the flux through a unit tube is 1, there will be $4\pi e$ unit tubes passing through the surface; or, with each unit of charge, there will be associated 4π unit tubes of force.

This convention is not the one commonly used, but it commends itself by being the same as that already used in the analogous case of the unit of magnetism. The ordinary convention assigns one unit tube of force to each unit of charge. On this convention the flux of force through a unit tube is given by $Rs = 4\pi$; and the relation between the force and the number of unit tubes that pass through unit area is given by

$$R = 4\pi N. \tag{137}$$

In demonstrating this relation we have assumed that the dielectric around the charged body is vacuum, or, for practical purposes, air; that is, that the dielectric constant of the medium is 1.

444. Tubes of Induction. — If a charged body is surrounded by a medium of which the dielectric constant is K, the electric force R at any point in the field is equal to P/K, where P represents the electric force that would be found at the same point if the medium had the dielectric constant 1. This can be seen at once by noticing that if the body is originally charged in vacuum so that its potential is V, the result of surrounding it by the dielectric is to reduce its potential to V/K; and since its potential is measured by the work done in moving a test unit from it to the surrounding conductors at potential 0, the force at each point by means of which the work is done must be reduced in the same ratio as the potential.

When we apply Gauss's theorem to this case, the total flux, measured by the electric force at each point of the closed surface enclosing the charge, becomes $4\pi e/K$.

If we construct tubes of force so drawn that $4\pi e$ tubes are still associated with a charge of strength e, the flux through each tube will be equal to $1/K$. These tubes will be duplicates in size and position of those which are drawn in the field in which the dielectric constant is 1, and differ from them merely in the magnitude of the flux in each one of them. The electric force R at any point is then equal to N/K, the number of such tubes of force which pass perpendicularly through unit area, divided by the dielectric constant.

The force in the field may be measured in another way. We conceive two surfaces drawn very close to one of the equipotential surfaces of the field and on either side of it, and we suppose the medium removed from the space between them, so that the dielectric constant within this space is 1. We then apply Gauss's theorem to the equipotential surface drawn in this space. The charge on the body will induce charges on the surfaces of the dielectric enclosed within the surface, one of which lies just outside the charged conductor, the other just inside the equipotential surface considered (Fig. 252); but these charges are known to be equal in amount and opposite in sign (§ 437), so that the total flux over the surface is $4\pi e$. Since, by our convention, $4\pi e$ unit tubes are associated with the charge e, the flux of force in each unit tube, measured within a narrow cavity or crevasse, formed in the way described, is equal to 1. Tubes of force in which the force is conceived to be measured in this way are called *tubes of electric induction*, or *tubes of electric flux*. If we designate the electric force, measured in this way, by P, and call it *the electric induction*, we have

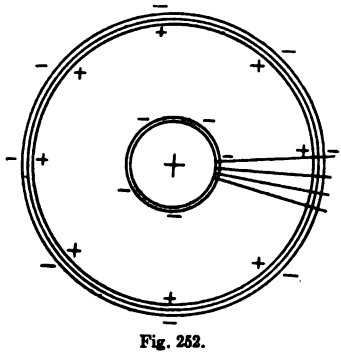


Fig. 252.

$$P = KR, \tag{138}$$

and may say that the electric induction is equal to the electric force multiplied by the dielectric constant.

The induction P is equal to N, the number of tubes of force, or better, of tubes of induction, which pass perpendicularly through unit area.

445. Charges at the Ends of a Tube of Force. — The lines of force which form a tube of force begin and end on charged conductors. Let us prolong them into the conductors and draw cross

sections s_1 and s_2 within the conductors parallel to their bounding surfaces (Fig. 253). We then have a closed surface formed of these end surfaces and the tubular surface bounded by lines of force. By Gauss's theorem the total electric flux through this surface is $4\pi\Sigma e$, where Σe is the total charge enclosed by the surface. But the total flux through the surface is zero, for the fluxes at the end surfaces are zero, since there is no electric force within a conductor, and the flux through the tubular surface is zero, since the force has no component perpendicular to the lines of force.

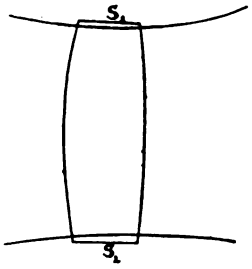


Fig. 253.

Hence $4\pi\Sigma e = 0$, or $\Sigma e = 0$,

and so if the charge on one of the conductors on the area s_1 determined by the tube of force is positive, the charge on the corresponding area s_2 on the other conductor is negative and equal to it in amount. Tubes of force therefore begin and end on equal and opposite charges.

This demonstration holds equally well when the field is filled with any dielectric.

446. Tension in Tubes of Force. — On Faraday's theory of medium action, the force acting on a charged body is due to stresses in the dielectric surrounding it. These stresses may be considered to be a tension equal to $R/8\pi$ along the axis of each unit tube of force and a pressure of equal amount in all directions perpendicular to that axis. We may calculate the magnitude of the tension from our knowledge of the force which acts on a charge on the surface of a conductor.

It has been shown in Examples, XXIII, 12, that a so-called electric pressure acts outwardly from the surface of a charged conductor at every point of it, equal to $2\pi\sigma^2$, where σ represents the surface density of the charge. This expression for the electric pressure was found on the supposition that the dielectric constant of the medium surrounding the conductor is 1. If it is equal to K, a proper modification of the demonstration (Examples, XXIII, 12) will show that the electric pressure is given by $2\pi\sigma^2/K$. We shall proceed with the discussion on this more general supposition.

The number of unit tubes of force connected with the charge σ on unit area is $4\pi\sigma$, so that the force due to the electric pressure acting outward on the end of each unit tube is $\frac{2\pi\sigma^2}{4\pi\sigma K} = \frac{\sigma}{2K}$. We can replace σ by the electric force R, for we know (§ 444) that the electric force just outside a conductor equals $4\pi\sigma/K$; so that we have finally for the force outward from the conductor on the charge at the end of each unit tube of force the expression $R/8\pi$.

If, therefore, we conceive that a tension acts along the axis of the tube of force equal to $R/8\pi$, such a tension will account for the force on the charge at its end, and the resultant force arising from all such tensions will be the force on the charged body.

A medium in which such a tension exists at every point will not in general be in equilibrium. That it shall be in equilibrium, a pressure equal to $R/8\pi$ must act in all directions perpendicular to the lines of force. The proof of this statement is so elaborate that it would be out of place here.

447. Energy in the Dielectric. — On the theory of action in a medium which we are now considering, the energy of a set of charged conductors is distributed throughout the dielectric surrounding them. We can find a possible mode of distribution as follows: It has been shown in Examples, XXIII, 13, that the energy of a set of charged conductors is given by $\frac{1}{2}\Sigma eV$. An examination of the demonstration will show that it does not depend on any properties of the dielectric in the field, and thus that the expression found for the energy will be valid in all cases. Let us consider the energy of the charges associated with any tube of force. These charges are equal and of opposite sign, so that, if the potentials of the conductors on which they reside are V_1 and V_2, the energy of the two charges is equal to $e(V_1 - V_2)/2$. All the charges may be associated thus in pairs, so that the total energy is equal to $\Sigma e(V_1 - V_2)/2$, where the summation is taken over the positive charges.

If now we assume that each unit length of a unit tube of force contains an amount of energy equal to $R/8\pi$, the energy of the charges will be accounted for. For, let d represent a small distance measured along the axis of a tube of force. Then $\Sigma Rd/8\pi$, if the summation is taken over the whole length of the tube, will represent the energy contained in a unit tube, and the $4\pi e$ unit tubes associated with the charge e, will contain $e\Sigma Rd/2$ units of energy. But $\Sigma Rd = V_1 - V_2$ by the definition of difference of potential (§ 429), so that the tubes associated with the charge e will contain $e(V_1 - V_2)/2$ units of energy, and all the tubes in the field will contain $\Sigma e(V_1 - V_2)/2$ units of energy. which is the energy possessed by the set of charged bodies.

If a unit area is drawn on an equipotential surface and the lines of force through its edge prolonged for unit distance, they will cut out another unit area on a neighboring equipotential surface, and the volume thus determined will be a unit volume. By taking the unit of measurement small enough such a unit volume may be determined, even when the lines of force are not parallel. By

our convention N unit tubes of force pass through the unit areas which are thus chosen, and each unit length of each tube contains $R/8\pi$ units of energy. The unit volume therefore contains $NR/8\pi$ units of energy. Since $N = P = KR$, the electric induction (§ 444), we also have the energy in unit volume given by

$$\frac{RP}{8\pi} = \frac{KR^2}{8\pi} = \frac{P^2}{K8\pi}. \tag{139}$$

448. Bodies in an Electric Field. — Within a conductor there is no electric energy; for the energy is measured by the force, and there is no electric force within a conductor. If a conductor is moved to any part of a field, the electric energy which previously existed there disappears. As it is the tendency of bodies to move so that the potential energy of the system becomes as small as possible, the spontaneous movement of a conductor will be from the weaker to the stronger part of a field. Furthermore, the tubes of

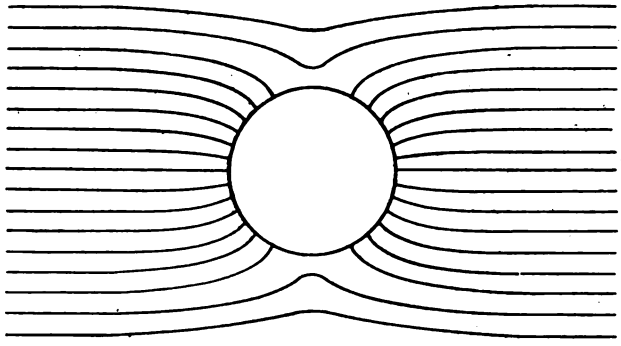


Fig. 254.

force of the field will be altered in shape, as if they were drawn into the conductor, so that the energy of the field is thereby reduced. Equilibrium is reached when the elongation of the tubes outside the conductor increases the energy at the same rate as the absorption of other tubes into it diminishes the energy. In a uniform field the tubes of force which meet the conductor terminate on negative charges, and an equal number of tubes originate on positive charges from the other parts of the conductor. The regions on which the negative and positive charges are found are separated by a line, called the neutral line, on which there is no charge. Fig. 254 represents the tubes of force ending on a spherical conductor in a uniform field.

There are no nonconducting substances known for which the dielectric constant K is less than 1. There are therefore no known phenomena in electricity corresponding to those of diamagnetism. The energy within a tube of force in a dielectric is always less than that in the same tube in vacuum. The tendency of any nonconducting body is therefore to move from a weaker to a stronger part of the electric field, and to draw the tubes of force into it.

449. Maxwell's Descriptive Theory of Electrification. — In connection with the theory of medium action, it is interesting to examine the particular form given it by Maxwell, rather as a description than as expressing any final theory of the real condition in a dielectric. Maxwell supposed the ether, and therefore all bodies in the ether, to be filled with electricity. We may best think of this electricity as existing in separate portions. Maxwell supposed that it moves freely, or with a resistance due only to friction, through conductors, but that its displacement in a dielectric is resisted by a force which increases with the extent of the displacement, and which he likened to elasticity. The process of charging a body then consists, according to this description, in a displacement of the electricity along the lines of force until it is checked by the electric elasticity. Considering the displacement with respect to the dielectric, it is inward when the dielectric is bounded by a positively charged conductor and outward when it is bounded by a negatively charged conductor. When connection is made by a conductor between two oppositely charged bodies, a flow of electricity passes through it and the strain in the dielectric is relieved. The work which is done in charging a conductor is, in this description, stored up in the dielectric as work done in effecting the electric displacement. To account for the various dielectric constants of different media, we suppose that the same force will produce different displacements in the different media.

EXAMPLES, XXIV

1. *To find the capacity of a freely electrified sphere.*

The charge on the sphere is e. The potential at the center of the sphere of radius R, and therefore the potential of the sphere in general, is e/R, so that the capacity is R.

2. *To find the capacity of a condenser consisting of two parallel plane plate conductors which are so near together that the lines of force between them are parallel straight lines.*

In enunciating the problem in this way, we neglect the charges on the edges of the plates.

The density of charge on the plate at potential V is σ. If the area of the plate is a, the charge on it is $a\sigma$. The electric force R near the plate at potential V is $R = 4\pi\sigma$, and since the lines of force are parallel, the force has the same value everywhere between the plates. If the distance between them is d, we have (§ 431) $R = V/d$, the second plate being supposed to be at zero potential, and hence $V = 4\pi\sigma d$.

The capacity $$C = \frac{e}{V} = \frac{a}{4\pi d}.$$

Even if the difference of potential V between the two plates is small, the charge e on a plate may be large if the distance d is small. If the plate at potential V is joined to an electroscope, the electroscope may show but slight indications while the two plates are near each other. If the plates are then separated, the capacity of the plate will be very much diminished, and the charge remaining the same, the potential will rise considerably, and the indications of the electroscope will be correspondingly increased. An electroscope arranged in this manner, and capable of indicating small differences of potential, was invented by Volta. It is called the *condensing electroscope.*

3. *To find the force exerted on the circular plane sheet of Examples, XXIII,* 11, *in terms of the difference of potential between the plates.*

From Example 2 we have $V = 4\pi\sigma d$, so that $\sigma = V/4\pi d$, and from Examples, XXIII, 11,

$$F = 2\pi a\sigma^2 = \frac{2\pi a V^2}{16\pi^2 d^2} = \frac{aV^2}{8\pi d^2}.$$

If F is measured in dynes, and a and d in centimeters, the formula $V = d\sqrt{\frac{8\pi F}{a}}$ gives the difference of potential in absolute electrostatic units. A formula derived from this one is used in determining difference of potential with the attracted disk electrometer (§ 436).

4. *To find the potential of a system of several conductors, previously electrified to different potentials, when they are joined to one another by thin conductors which do not change the capacity.*

The charges on the conductors are C_1V_1, C_2V_2, . . . C_nV_n. After connections are made these charges are distributed over the system, and their common potential is V. The capacity of the system is the sum $C_1 + C_2 + \cdots C_n$ of the capacities of the conductors. We have therefore

$$V = \frac{C_1V_1 + C_2V_2 + \cdots C_nV_n}{C_1 + C_2 + \cdots C_n}.$$

When C_n is very great in comparison with the other capacities, $V = V_n$ approximately.

The earth is a body whose capacity is very large in comparison with that of ordinary conductors, so that when other conductors are joined to earth, they assume the potential of the earth, which remains practically constant. This circumstance justifies the use of the potential of the earth as a standard or zero potential.

5. *To find the capacity of a condenser similar to that of Example* 2, *when a medium of dielectric constant K fills the space between the plates.*

The difference of potential remaining the same as in Example 2 the capacity becomes K times the capacity obtained in that example, or is $C' = KC = Ka/4\pi d$.

6. *To find a distribution of electricity which will render the electric force, measured in a disk-shaped cavity or crevasse taken in the dielectric with its faces normal to the lines of force, equal to the number of tubes of force that pass perpendicularly through unit area in the dielectric.*

The electric force $R = N/K$ in the dielectric. This force acts in the crevasse, and that the whole force in the crevasse shall be $P = N$ we must have $P = R + N - R = N/K + N(K - 1)/K$. If the faces of the crevasse are covered with equal uniform distributions of positive and negative electricity of surface density $\sigma = N(K - 1)/4\pi K$, the necessary additional force $4\pi\sigma = N(K - 1)/K$ will act in the cavity and the total force P will equal N. The force measured in this way is the *electric induction*.

CHAPTER XX

THE ELECTRIC CURRENT

450. Galvani's Discovery. — In 1791 the Italian physiologist Galvani happened to notice that the legs of a recently killed frog, lying near an electric machine, were thrown into convulsions whenever a spark passed from the machine. He investigated this phenomenon on the hypothesis that the nerves and muscles act like the coatings of a charged Leyden jar, and in following up this hypothesis he found that the convulsive movements occurred whenever the lumbar nerves and the muscles of the leg were joined by a metallic connection. The action, which was comparatively slight when this connection was made by one metal, was made much greater by touching the nerve with a strip of one kind of metal, the muscles with a strip of another kind, and then bringing the two strips together. Galvani interpreted this result consistently with the hypothesis already stated, but other observers were led to consider the frog's legs simply as a very sensitive electroscope, and to ascribe the action observed to the contact of the metals.

451. Contact Difference of Potential. — In the years 1798–1802 the Italian physicist Volta succeeded in demonstrating the production of electrification by the contact of metals, and in applying his discovery to the construction of an apparatus for the production of the electric current. By the use of the condensing electroscope (Examples, XXIV, 2), which he had invented a few years before, Volta was able to prove that when pieces of two different metals, like copper and zinc, are brought in contact, the electric condition of the one becomes different from that of the other. As we should now describe it, the potentials of the two metals become different from each other when the metals are brought in contact. This difference of potential is very slight, altogether too slight to be detected by an observation with any but a sensitive electroscope.

452. Volta's Series. Conductors of the First Class. — Volta demonstrated that the difference of potential occurring on contact is a definite characteristic difference for each pair of metals employed. He proved also that the electric effect produced by arranging a number of different metals in contact with one another

in succession is to charge the two metals at the end of the row, just as they would be charged if they were immediately in contact. By measuring the different potential differences existing between the different pairs of metals, Volta found that the algebraic sum of the potential differences arising at the successive contacts is equal to the difference of potential arising from the contact of the two terminal metals of the arrangement, so that if a closed series of different metals is formed, by bringing the two terminal metals in contact, the algebraic sum of the differences of potential arising from the contacts will equal zero. That this is the case appears also from the fact that there is no evidence of any continuous movement of electricity, arising from an unbalanced potential difference, in such a metallic circuit. Taking some one metal as a standard, and determining the potential differences arising from its contact with other metals, we obtain data from which the difference of potential arising from the contact of any two of the metals may be determined. Such a set of data is called a *Volta's series*, or the electromotive series. Volta used the term, *conductors of the first class*, to designate those substances which give rise to such potential differences that their algebraic sum equals zero, when the substances are arranged in a closed circuit.

When two metals are in contact, the one of them which shows a positive potential is said to be electropositive to the other. The one which shows a negative potential is electronegative. The Volta's series begins with the most electropositive metal, which becomes positive on contact with any one that follows it in the series, and ends with the most electronegative metal, which becomes negative on contact with any one that precedes it in the series. Any intermediate metal becomes positive with those that follow it, and negative with those that precede it. The following list contains a few of the most important elements arranged in the electromotive series:

\+ zinc, tin, lead, mercury, iron, copper,
silver, carbon, platinum −.

453. Conductors of the Second Class. — Volta could not discover any difference of potential arising from the contact of a metal with a liquid. He therefore called the liquids *conductors of the second class*. We now know that potential differences do arise from such a contact, but that they are neither of such a magnitude nor of such a sign as to reduce the sum of the potential differences

in a circuit to zero, when a conductor of the second class forms a part of the circuit. In a typical circuit of this sort, in which two different conductors of the first class are in contact with a conductor of the second class, there exists an unbalanced difference of potential. We shall see later that conductors of the second class are always substances which act chemically upon the conductors of the first class, and that a liquid like mercury, by which no such chemical action is exerted, is in the first class.

454. The Voltaic Battery. — Following out the indications of his theory, Volta undertook to increase the potential difference developed by contact by bringing together a number of successive pairs of the same metals. To do this he made a number of disks of zinc and an equal number of disks of silver. A disk of zinc placed on a disk of silver formed what we may call an element. He then piled a number of these elements one above the other, separating the zinc of each element from the silver of the one above it by a disk of paper or felt moistened with water, in which, to render it a better conductor, salt or acid was dissolved. This arrangement we now call the voltaic pile (Fig. 255). Conductors were joined to the disks of zinc and silver at the two ends of the pile, and their electric condition was observed. It was found that the pile was charged oppositely at its two ends, and exhibited the phenomena observed with static electricity. The pile behaved, as Volta said, like a battery of Leyden jars, with this difference, that the jars, when discharged, must be charged from an outside source before another discharge can be obtained from them, whereas the pile charges itself, so that the electric effects can be obtained from it continuously.

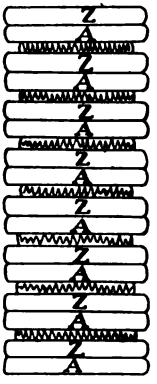


Fig. 255.

The difference of potential between the conductors joined to the two ends of the pile is approximately equal to the difference of potential due to the contact of silver and zinc, multiplied by the number of elements in the pile. According to Volta's interpretation of the facts, a difference of potential arises in each element from the contact of the two metals, and the potential of the two metals in neighboring elements is the same, by reason of the presence between them of a conductor of the second class. Thus each successive element introduces an additional potential difference.

Volta soon recognized that a pile of this sort is not so well adapted for continuous use as another arrangement of the conductors. He

therefore constructed the arrangement which he called the crown of cups, and which we now call the *voltaic battery*. Each element of this battery is called a *voltaic cell*. As now constructed a typical voltaic cell consists of a glass vessel partly filled with acidulated water, in which are immersed a plate of copper and a plate of zinc (Fig. 256). Wires are attached to the upper ends of these plates, by which the difference of potential existing between them can be transferred to any desired point. The voltaic battery consists of a number of such cells, of which the zinc of one is joined by a wire to the copper of the next.

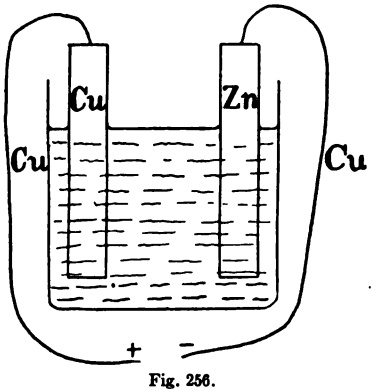


Fig. 256.

In the typical case of copper and zinc and of the cell formed with them it may be interesting to notice that when zinc and copper are brought in contact, the potential of the zinc becomes positive to that of the copper. The zinc is therefore *electropositive* to copper. When the zinc and copper of the cell are furnished with wires of the same sort, the potential of the wire joined to the copper is positive as compared with the potential of the wire joined to the zinc. As we commonly think of the electric current as flowing from the higher to the lower potential in that part of the circuit which lies outside the cell, we call the copper plate the *positive pole* of the cell, and the zinc plate the *negative pole*.

455. Heating Action of the Current. — Volta considered the voltaic battery as a source of continuous electric discharge. It was this conception of the action in the circuit which led to its being called the *electric current*. Volta found that the pile would give the shock which is felt in the body when a Leyden jar is discharged through it. He also found that when the terminal poles of the pile, or of the battery, were joined by a thin wire, the wire became heated. A similar heating effect had been observed when a battery of Leyden jars was discharged through a wire. By the use of modified forms of the battery, a very great amount of heat was developed in the circuit, and it was shown that the heat developed depends in some way upon the nature of the materials composing the circuit. The quantitative law of the development of heat was discovered by Joule in 1842.

456. Chemical Action of the Current. — When Volta's account of the battery reached England in 1800, two members of the Royal Society, Nicholson and Carlisle, constructed a battery and made

experiments with it. In one of these they introduced the ends of the terminal wires into a drop of water, and noticed that bubbles of gas arose from them. On investigating this phenomenon further, they found that when the terminal wires were of platinum and were separated by a column of water, different gases were evolved at the two wires. One of these gases was recognized as hydrogen.

Other contemporary students investigated this action with more success by using an apparatus constructed as follows (Fig. 257): Two test tubes filled with acidulated water were placed with the open ends down in acidulated water contained in a vessel, the water being held up in the tubes by atmospheric pressure. In the open ends of the tubes were placed small platinum plates, to which were joined the terminal wires of the battery, these wires being covered with insulating material, so that the current could not enter the water except from the platinum plates. When the circuit was completed, gases were evolved at both the platinum terminals, and collected in the upper ends of the tubes. These gases, when examined, were found to be oxygen and hydrogen, the constituents of water. It appeared from this experiment that the electric current can decompose water into its constituents. When the indications of this experiment were followed up, it was found that very many compounds undergo a similar decomposition. A common characteristic of all the compounds which undergo decomposition was found to be that the compound must be brought into the liquid state, either by solution in a solvent or by fusion. The products of decomposition which are obtained are not always the constituents of the compound which is dissolved. In the case just described, for example, the products obtained are not the constituents of the acid dissolved in the water, but of the water itself. In such cases the direct action of the current is supposed to be complicated by secondary chemical actions.

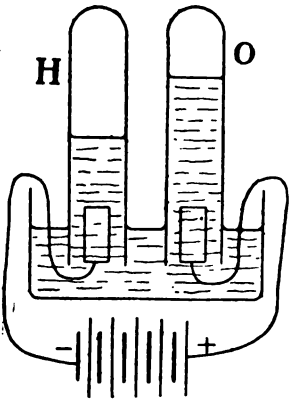


Fig. 257.

The quantitative laws of the chemical action of the current were discovered in 1834 by Faraday.

457. The Electric Arc. — In 1800, by the use of a powerful battery, Davy discovered that when the two terminal wires, or better, when two pieces of charcoal or carbon which are joined to the

terminal wires, are first touched together, so that the current is established, and are then slightly separated, the ends near the point of contact become intensely heated and brilliantly luminous. At the same time a luminous flamelike column, or the *electric arc*, appears between them. The light in the arc itself is generally bluish in color, and is less intense than that from the carbon terminals. When carbon rods are used as terminals, a small depression, called the crater, is formed at the end of the positive terminal. The highest temperature of the arc is obtained in this crater. It is estimated to be as high as 3400° C. All known substances except carbon can be fused in the arc.

While the arc is established, both the carbon rods waste away, the positive rod wasting twice as fast as the negative rod. This effect is not due entirely to combustion, for the arc may be established in a vacuum, and the wasting away of the terminals occurs in that case also.

Electromagnetism

458. The Magnetic Field of the Current. — The first substantial advance which was made in the knowledge of the properties of the electric current was based upon the discovery that a magnetic field is associated with the current. This discovery was made in the year 1820, by the Danish physicist Oersted, who had been for some time hoping to find a relation between the current and magnetism. He placed a wire carrying a current above a compass needle and parallel to it, and noticed that the needle turned out of the magnetic meridian. When the current was reversed, the needle turned out of the meridian in the other sense. When the wire was placed below the needle, its deflection for the same direction of the current was opposite to that obtained when the wire was above it.

Oersted's discovery opened the way to the determination of the laws connecting the current with its magnetic field, and to the development of the science of electromagnetism.

459. Biot and Savart's Experiments. — As soon as Oersted's discovery become known, Biot and Savart undertook a quantitative examination of the action of a current on a magnet. They investigated only the simplest case, in which the current is straight and practically infinitely long. A long straight wire was set up vertically and magnets were arranged about it so as to neutralize

the earth's magnetic field in its neighborhood. A region was thus obtained about the wire in which there was no perceptible magnetic force, so long as there was no current in the wire. A small magnet, suspended so as to turn with freedom in any direction, was used as an indicator of the effect of the current. When the current was set up, this magnet assumed a position in which it was tangent to a circle drawn around the wire as center, in a plane perpendicular to it (Figs. 258, a, b). When the current was in one direction, the north pole of the magnet always pointed in one sense around this circle, at whatever point on the circle it was placed. When the current was reversed, the direction in which the magnet pointed was also reversed. We may express this result by saying that the current sets up and maintains a magnetic field, in which the lines of force, in the case of a long straight current, are circles having the current as a common center.

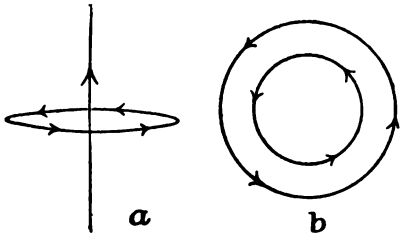


Fig. 258.

460. **Direction of the Magnetic Force.** — In order to express the relation between the direction of the current and the direction of the lines of force conventions must be made to determine these directions. The current is assumed to flow from a point of higher potential to a point of lower potential in the circuit containing a voltaic cell, or from the positive pole of the cell through the external circuit to the negative pole. The direction of a line of force is

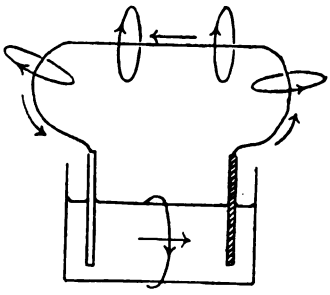


Fig. 259, a.

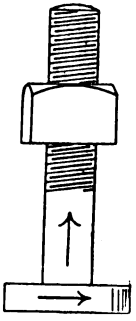


Fig. 259, b.

Fig. 259, c.

(§ 390) the direction in which the force of the field would urge a free north pole to move, or the direction in which the north pole

of a freely suspended short magnet would point. With these conventions we may express the relation between the direction of the current and the direction of the force in its magnetic field by using any one of several modes of statement.

Ampère's rule is that when an observer swimming with the current looks toward the magnet, the north pole of the magnet is deflected toward his left. Maxwell's rule is that the direction of the current and the direction of the lines of force are related as the translation and the rotation of a right-handed screw (Figs. 259, a, b). Another rule is that if the right hand grasps the wire, with the thumb extended in the direction of the current, the fingers encircle the wire in the direction of the lines of force (Fig. 259, c).

461. Law of the Magnetic Force. — By allowing the indicator magnet to execute vibrations in the magnetic field of the current, at different distances from the wire, Biot and Savart showed that the magnetic intensity at any point in the field was inversely as the distance of that point from the wire. If we use k to represent a factor of proportion, and p to represent the distance from any point to the wire, the magnetic intensity R at that point may be expressed by the formula $R = k/p$.

The factor k is not an absolute constant, but depends upon the character and arrangement of the parts of the circuit, especially upon the number and disposition of the voltaic cells used to set up the current. We may say in general that it depends upon the strength of the current. In default of any other way of measuring the strength of the current we may define it for the present by setting it proportional to k, so that if i represents the strength of the current and n a new factor of proportion, we have $k = ni$. With this definition of strength of current the previous formula becomes $R = ni/p$. The statement that the magnetic intensity at any point near an infinitely long straight current is proportional directly to the strength of the current, and inversely to the distance of the point from the current, is often called *Biot and Savart's law.*

The definition here given of strength of current is consistent with the other definitions which will finally be adopted. It is not expedient to define a unit current from this relation by setting $n = 1$, as more satisfactory definitions will be obtained hereafter from other considerations.

Laplace showed that Biot and Savart's law, which expresses the integral effect of the whole straight current, can be obtained from the hypothesis that each element of the current of length ds (Fig. 260) acts on the unit magnetic

pole with which the field is tested, with a force which is proportional directly to the strength of the current, and inversely to the square of the distance r between the element and the pole, and which is also proportional to the length ds of the element and to the sine of the angle θ between the line joining the element with the pole and the direction of the current in the element. This elementary force acts perpendicularly to the plane containing the current and the pole, and the total force or magnetic intensity R, acting downward through the paper, is the sum of all the elementary forces. The formula embodying Laplace's rule may be written with the help of a factor of proportion c in the form $R = c\Sigma\frac{i \sin\theta\, ds}{r^2}$. When we measure the current in the electromagnetic unit of current, which will hereafter be defined, we set $c = 1$. In this case the constant n of Biot and Savart's formula is equal to 2 (Examples, XXV, 2).

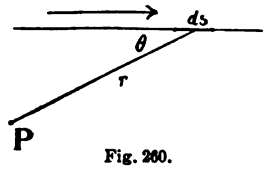


Fig. 260.

It should be noticed that the force which is exerted on a magnetic pole in the neighborhood of a long straight current has its counterpart in an equal force which acts on the conductor carrying the circuit, in a line parallel with the force acting on the pole and in a direction opposite to it. The two forces constitute a couple. We escape from the difficulty presented by the fact that the forces thus considered do not conform to the third law of motion by recognizing that the straight conductor does not constitute the whole of the circuit from which action upon the pole arises, and that separate magnetic poles do not exist. When the force exerted by a magnet on a closed circuit is considered, it is opposite to the force on the magnet, as well as equal to it.

462. Forces between Currents. — The fact that a current acts on a magnet, by means of the magnetic field which is associated with it, suggested that two currents might perhaps interact with each other by means of their magnetic fields. By experiments instituted to test this suggestion, Ampère (1826) showed that in fact currents do interact with each other. He furthermore determined a formula or law expressing this interaction, on the supposition that each element of current acts on every other element of current with a force which conforms to the third law of motion, and which depends upon the directions of the elements, the direction of the line joining them, and is inversely proportional to the square of the distance between them. With this formula all the forces exerted by currents on each other may be calculated.

So long as we confine our attention to limited portions of two circuits carrying currents, and consider only the forces exerted by those portions on each other, we may express the general mode of action discovered by Ampère by saying, that parallel currents which are in the same sense attract each other and those which are in opposite senses repel each other. If the currents are not parallel, the forces are such as would be exerted by parallel com-

ponents of the two currents, and are therefore attractions if the components are in the same sense, and repulsions if they are in opposite senses.

The formulas of Laplace and Ampère, which express the actions of currents on magnets and of currents on each other as the result of direct actions radiating from each element of current, were in accordance with the thought of the time in which they were proposed. Speculative physics had been powerfully influenced, first by the discovery of the law of gravitation, and then by the discovery that similar laws held for the forces between magnets and between electric charges. These laws express the forces observed as the integral or sum of elementary forces acting between elementary masses, magnetic poles, or electric charges in the line joining the interacting elements and varying inversely as the square of the distance between them. They naturally served as the model for the analysis of the newly observed forces due to electric currents, and the reduction of these forces to forces between current elements or between current elements and a magnetic pole, which conform, though with additional complications, to a similar law, was received as a satisfactory explanation of them.

It has already been shown (§§ 398, 446) that an alternative explanation of magnetic and electric forces can be given by ascribing them to a peculiar condition of a universal medium, in which tensions exist in the medium between the magnetic poles or electric charges. We shall show that the explanation of the magnetic field in this way can be extended to the case of the magnetic field set up by an electric current, and that the magnetic relations of the current can be adequately described by the aid of this theory of medium action. Ampère, quite unconsciously, prepared the way for the development of this later explanation by his discovery of a most interesting similarity between the magnetic field around a current and the magnetic field of a magnetic shell.

463. The Equivalent Magnetic Shell. — By a direct calculation from his formula for the interaction of current elements on each other, Ampère showed that the force between two closed circuits carrying currents is similar to the force between two magnetic shells (Examples, XXI, 5), if the strengths of the magnetic shells are uniform and if their edges coincide with the two circuits. By a suitable adjustment of the strengths of the shells the forces may be made not only similar but equal. Furthermore, the force between a closed circuit carrying a current and a magnetic pole at any point in the field is proportional to the force between a magnetic shell and the magnetic pole, if the strength of the magnetic shell is uniform and if its edge coincides with the circuit. In this case also, the force exerted on the pole by the shell may be made equal to that exerted by the current, by a suitable adjustment of the strength of the shell. Such a magnetic shell, which will set up a magnetic field similar to that of a closed current, is said to be *equivalent* to the current.

Since the direction of the magnetic force associated with a current depends on the direction of the current, it is necessary for us to consider the relation between the direction of the current in the circuit and the distribution of magnetism on the equivalent magnetic shell. The application of any of the rules already given (§ 460) shows at once that the forces due to the current and those due to the shell are in the same direction when, to an observer looking at the south face of the shell, and so along its lines of force in their positive direction, the current is traveling clockwise in the equivalent circuit (Figs. 261, a, b).

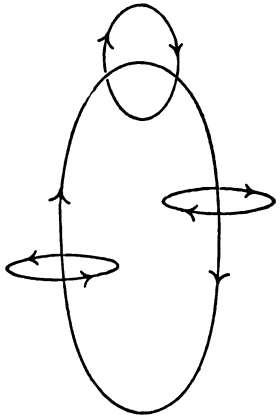


Fig. 261, a.

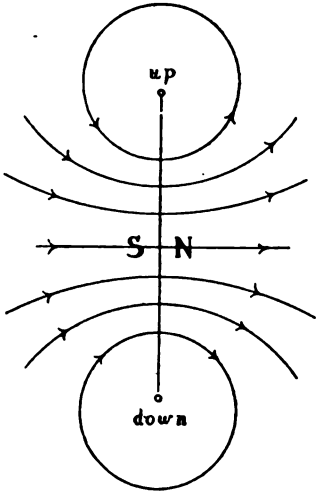


Fig. 261, b.

The equivalence between the field of a magnetic shell and the field of a current may be shown in a more immediately experimental way by a method due to Weber. Weber showed by direct experiment that the field due to a short bar magnet and the field due to the current in a small circuit were entirely similar to each other at any distance from either which is great in comparison with their dimensions; and it is a reasonable inference that the fields would be similar at all distances if the dimensions of the magnet and the circuit could be taken small enough. Now by laying many similar short bar magnets side by side, with their north poles all pointing in one direction, we can build up a magnetic shell (Fig. 262, a) and by arranging side by side many small circuits carrying equal currents, we can build up the equivalent of an electric current circulating in the periphery of the arrangement (Fig. 262, b). The currents in the contiguous small circuits will be

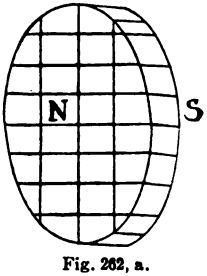


Fig. 262, a.

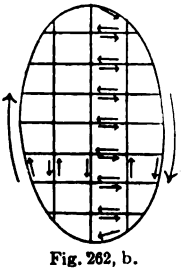


Fig. 262, b.

oppositely directed and will cancel each other as currents, leaving only those currents uncanceled which are flowing, and all in the same sense, in the periphery. When the periphery of the arrangement of circuits is of the same size and shape as the edge of the magnetic shell, the two fields will be similar.

Since the fields of the current and of its equivalent magnetic shell are similar, the potential at any corresponding points in the two fields will be the same. The potential at a point in the field of a shell is $j\Omega$ (Examples, XXI, 5). Anticipating the choice of a unit current which will be made in § 465, we may represent the potential at a point in the field of a current of strength i by $i\Omega$. In these formulas Ω represents the solid angle subtended by the shell or by the circuit at the point at which the potential is desired.

464. Multiple-valued Potential of the Current. — The field of the current differs from that of the equivalent magnetic shell in that the lines of force in the field of the shell proceed from points on its north face to corresponding points on its south face, so that they are interrupted by the material of the shell, while the lines of force of the current are closed continuous lines encircling the current. Furthermore, the lines of force in the interior of the shell are not similar to those in the corresponding part of the field of the current. This second point of difference can be removed by measuring and tracing the magnetic induction of the shell instead of the magnetic force or intensity. The lines of induction are closed continuous curves, and are the precise counterpart of the lines of force of the current.

The first point of difference gives rise to a most important contrast between the field of the shell and the field of the current, which may best be studied by an examination of the magnetic potentials in the two fields.

It may be shown (Examples, XXI, 5) that the potential at a point in the field of a magnetic shell of strength j is expressed by $j\Omega$, where Ω represents the solid angle subtended by the shell from the point. The potential in this case may be thought of as arising from the combined action of the elementary quantities of magnetism distributed on the faces of the shell, and is like the potential in all other cases in which the field is due to the action of centers of force, in that it depends solely upon the position of the point in the field, and is independent of the path by which the unit pole used to measure it is removed to infinity. It also possesses the property of other fields of a similar character, in that the work done on a unit pole which is carried from a point around any path and brought back to its starting point is zero. If, for example, the unit pole is placed at a point on the north face of the shell, where its potential is $2\pi j$ (Examples, XXI, 6), and is moved through any path to a point on the south face, where its potential is $-2\pi j$, so that the field has done $4\pi j$ units of work on the pole, it is possible to bring the pole back to the starting point on the positive face only by carrying it around the edge of the shell; and if this is done the work done on the pole by the forces of the field is $-4\pi j$, so that the work done in the closed path is zero.

On the other hand, if a similar experiment were to be tried with the current of strength i, which will set up a potential $i\Omega$ at a point in the field, the work done in carrying the pole from the north face of the circuit to its south face around the outside of the circuit is $4\pi i$, and to bring the pole back to its starting point

it is not necessary to carry it around the outside of the circuit, but it may simply be moved an infinitesimal distance through the plane of the circuit along a line of force. The force in the field is everywhere finite, and the work done in this infinitesimal displacement is infinitesimal, so that the pole may be brought back to its starting point in such a way that work equal to $4\pi i$ has been done upon it. In general this result is accomplished by carrying the pole over a path which encloses the current.

The potential at a point in the field of a current, as measured by the work which is done on a unit pole as it moves from the point to an infinite distance, is therefore not a definite magnitude $i\Omega$, but may be made to differ from this by $4\pi i$, or by any multiple of $4\pi i$, by carrying the pole to an infinite distance over a path which encloses the current once, or as many times as we please. If n represents the number of times the path encloses the current we may write the potential at a point, $V = i(\Omega + 4\pi n)$. The number n may have any value, from 0 to ∞. The potential in such a field is said to be multiple-valued, in contrast to the single-valued potentials of the field arising from central forces.

465. The Electromagnetic Unit of Current. — The complete equivalence which may be established between the magnetic field of a current in a closed circuit and the magnetic field of a magnetic shell, by a suitable adjustment of the strength of the shell, makes it possible for us to define a unit of current which may be used in all our study of the relations of currents to magnetic fields. To do this we consider a circuit carrying a current, and its equivalent magnetic shell. We then define the unit of current as being a current such that the number of units of current in the circuit and the strength of its equivalent magnetic shell are equal. Or otherwise, if we consider a magnetic shell of unit strength, whose edge coincides with a closed circuit, then the current in that circuit which will set up a magnetic field identical with that of the shell is the unit current.

The current thus defined is called the electromagnetic unit of current. It is found convenient in practice to use a current as unit which is equal to 10^{-1} electromagnetic units. This practical unit is called the *ampere*.

466. Unit of Quantity. — The electromagnetic unit of quantity of electricity or electric charge is derived so immediately from this definition, combined with the conception that the current is electricity transferred through the conductor, that it may properly be considered here. It is defined as the quantity which will pass through any cross section of a conductor in one second, if a current of one electromagnetic unit exists in the conductor. The practical unit of quantity is the quantity which will pass through any cross section of the conductor in one second if a current of one ampere

exists in the conductor. This practical unit is called the *coulomb* (§ 424).

467. Ampère's Theory of Magnetism. — Ampère used the equivalence between a current and a magnetic shell as a basis for a theory of magnetism. He supposed each molecule of iron to be a separate and permanent magnet. He then explained the magnetic condition of the iron molecules by supposing that each of them has an electric current continually circulating about it, which sets up a magnetic field similar to that around a very small magnet. As in the later theory of Weber, the magnetism of the magnet as a whole is accounted for by a suitable arrangement of the molecular magnets. This theory of magnetism, or one that is essentially similar to it, though expressed in other terms, is still the most satisfactory way of explaining natural magnetism.

468. Electromagnets. — If a long wire is coiled into a tight spiral, each turn is practically a small closed circuit and acts like a magnetic shell, so that the magnetic field of the whole spiral is like that of a pile of magnetic shells arranged with their similar faces in the same direction. Such a pile of magnetic shells is plainly equivalent to a magnetized bar, and the magnetic field of such a bar is similar to that of the spiral. The spiral used in this way to replace a magnet was called by Ampère a *solenoid*.

An important difference between the field of the solenoid and that of the magnet must be especially noticed. The lines of force of a magnet, so far as they can be traced by experiment, run from the north end of the magnet to its south end. On the other hand, the lines of force of the solenoid are closed curves, any one of which may be traced from a point at one end of the solenoid through the region outside it to the other end, and then on in the same sense within the solenoid to the point of beginning. The lines of force of the solenoid therefore form a bundle of closed curves.

If instead of examining the magnetic force of the magnet which is equivalent to the solenoid, we examine the magnetic induction, and draw the tubes of induction (§ 405), we find that they form closed tubes which are in every respect similar to the tubes of force of the solenoid. We may therefore express the relation between the solenoid and a magnet by saying that the tubes of force of the solenoid are similar to the tubes of induction of the equivalent magnet. For a similar reason we may say in general that the tubes of force of a current are similar to the tubes of induction of the equivalent magnetic shell.

If a bar of soft iron is placed within the solenoid, while it carries a current, the magnetic field of the solenoid will make the bar a magnet. Such an arrangement is called an *electromagnet.* The magnetization of the bar disappears, or nearly so, when the current in the solenoid ceases. Its intensity of magnetization depends upon the quality of the iron, but mainly upon the strength of the magnetic field due to the current, and this depends on the strength of the current and on the number of turns made by the circuit about its axis in a unit of length.

469. Galvanometers. — Any instrument used to detect the presence of a current, or to measure its strength by observations of the interaction between the current and a magnet, is called a *galvanometer.* A simple circuit of wire placed vertically in the plane of the magnetic meridian, with a magnet suspended in the middle of it, will answer this purpose. The current flowing in the circuit sets up a magnetic field of which the lines of force at all points in the plane of the circuit are perpendicular to the magnetic meridian, and the forces of this field turn the magnet out of the magnetic meridian.

The effect is naturally heightened when the wire is turned on itself many times into a flat spiral, the strength of the magnetic field increasing with the number of turns. In one of the early forms of galvanometer, constructed by Schweiger, a coil or circuit of this sort was used in combination with an astatic needle (Fig. 263). The *astatic needle,* or system, consists of two similar light magnets held rigidly parallel to each other by a short connecting rod. These magnets are magnetized in opposite senses, and one of them a little more strongly than the other. When such a pair of magnets is suspended, the stronger one will overpower the other, and its north pole will point toward the north, but the directive action of the pair is much feebler than that of either one of them. In Schweiger's instrument the lower magnet of the two hangs within the coil, while the other one is above it. When the current passes, its magnetic field turns both these magnets in the same sense, and since the directive action of the magnets in the field of the earth is slight, the deviation of the system will be great in comparison with that which the same current would produce in a single magnet. Very feeble currents may therefore be detected by this instrument.

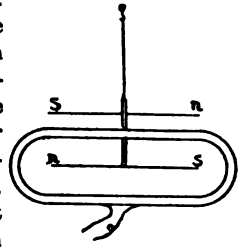


Fig. 263.

A galvanometer which is of special practical and theoretical importance is the *tangent galvanometer,* so called from one of its characteristic properties (Figs. 264, a, b). The coil of this instrument consists of a number of turns of wire wound in a circle, so that the thickness of the coil is very small in comparison with the

radius of the circle. This circle is set up on edge in the plane of the magnetic meridian, and a short magnet is suspended at its center. When a current is sent through the coil, the magnet is turned out of the meridian, and assumes a position in which the couple exerted on it by the earth's field is in equilibrium with the couple exerted on it by the field of the coil. The study of the field of the coil shows that it is of constant intensity in a region lying around the center of the coil, so that the force exerted by the field of the coil on the magnet's pole does not change with the deflection of the magnet. If we represent the angle of deflection by ϕ (Fig. 265), the horizontal intensity of the earth's field by H, and the strength of the field of the current by R, and remember that the lines of force of the current are at right angles to the plane of the coil and so also at right angles to the magnetic meridian, it is evident that the couple exerted on the magnet by the earth's field is proportional to $H\sin\phi$ and the couple exerted by the magnetic field of the current is proportional to $R\cos\phi$, so that we have $H\sin\phi = R\cos\phi$, or

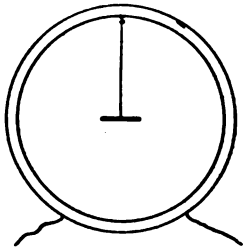


Fig. 264, a.

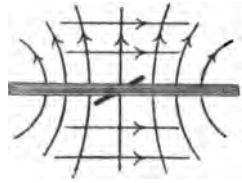


Fig. 264, b.

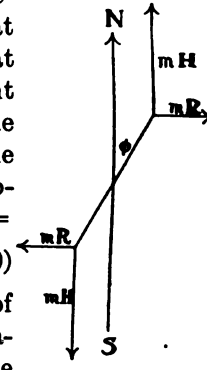


Fig. 265.

$$R = H\tan\phi. \qquad (140)$$

The ratio of R to H is the tangent of the angle of deflection. It is for this reason that this galvanometer is called the tangent galvanometer. If we make the supposition that the strength of the current in the coil is proportional to the strength of the magnetic field which it sets up, we may compare different currents by comparing the tangents of the deflections which they occasion.

The tangent galvanometer may be used to measure a current in terms of the electromagnetic unit (§ 465). It is easy to show (Examples, XXV, 1) that the magnetic intensity R at the center of a circular coil made up of n circles of approximately equal radius r, and carrying a current of strength i, is equal to $2\pi ni/r$. The horizontal intensity of the earth's field is known by previous observations (Examples, XX, 7). An observation of the deflection of the magnet will give the angle ϕ, and Equation 140 becomes $i = \frac{rH}{2\pi n}\tan\phi$, in which the strength of the current is expressed in terms of known quantities.

An examination of the formula by which the horizontal intensity H is expressed will show that it contains only quantities which are measured in terms of the fundamental units of mass, length, and time. The current, therefore, when measured by means of the formula just given, is also measured in terms of these fundamental units, and the electromagnetic unit of current is measured also in terms of these units. The unit thus measured is therefore an *absolute unit.*

470. Electromagnetic Rotations. — A number of arrangements were constructed by Faraday in 1821 by which the magnetic force acting between a current and a magnet was made to produce a continuous rotation of a part of the apparatus.

In a typical one of these instruments (Fig. 266) two shallow circular troughs are placed concentric with each other around an upright post, on the top of which is fixed a bearing. Two similar magnets are placed vertically on opposite sides of the apparatus, so that similar poles stand each between the troughs, and a little above their level, while the other poles are considerably below that level. A light wire frame, consisting of a horizontal portion, of one dependent portion which reaches the inner trough, and of two dependent portions which reach the outer trough, is mounted on a rod whose end rests in the bearing on the top of the post, so that the frame can turn about the rod as a vertical axis. Mercury is poured into the troughs until the ends of the dependent wires are always in contact with it as the frame turns round its axis. Wires from a voltaic battery introduce the current into one trough, from which it passes through the frame, and so out through the other trough. When the current is introduced, the frame maintains a continuous rotation about its axis. This movement may be understood by considering the force between one of the poles and that part of the movable circuit which is near it. If the circuit were at rest and the pole were free to move, it would be attracted toward one face of the circuit, would pass through it, and then be repelled by the other face. That is, the pole would move as nearly as possible along the lines of magnetic force of the current. Since the magnet is fixed, it is the circuit which moves in the opposite direction to that in which the magnet would move in the case supposed. If the circuit were everywhere solid, it would not be possible for the pole to pass through its plane without carrying the other pole with it, and the effect of the two poles being opposite, no continuous rotation would be set up. By making a part of the circuit fluid, we have arranged it so that the solid part of the circuit can pass from one side to the other of one pole without being affected by the opposite pole.

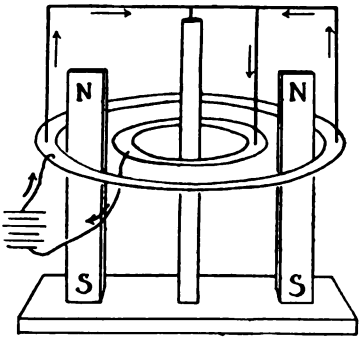


Fig. 266.

The work which is done in sustaining these motions comes from energy supplied to the circuit by the battery. As we shall afterwards see more at length, the circuit when in motion is not heated

so much as it is when it is at rest. The work which is done during the motion is the equivalent of the heat which has disappeared.

471. Energy of a Circuit Carrying a Current. — Since we have been able to represent the energy of a set of magnets by a distribution of energy in the medium occupying the magnetic field (§ 399), and since a circuit carrying a current sets up a magnetic field around itself, it ought to be possible to find an expression for the energy of the current which is distributed in its magnetic field. We shall now endeavor to do this, using the result of our previous investigation, that a unit length of a unit tube of induction, in a part of the field where the magnetic force is R, contains $R/8\pi$ units of energy.

When a circuit carries a current of strength i, the number of tubes of induction which it sets up and which pass through the circuit is proportional to the current strength, and by introducing a factor of proportion L, it may be set equal to Li. If we consider the energy in one of these tubes of induction we may find it by multiplying the energy in unit length of it by an element of length d, and adding these products for the whole length of the tube. The energy in the tube is therefore $\Sigma Rd/8\pi$. But ΣRd is equal to the work done on a unit pole as it traverses the tube, and this (§ 464) is equal to $4\pi i$, so that the energy in the tube is given by $i/2$, and is the same for each of the Li tubes. The whole energy in the field is therefore equal to $Li^2/2$.

The factor of proportion L is called the *coefficient of self-induction*. It is independent of the current, and depends upon the size and shape of the circuit, and the nature of the surrounding medium.

472. Energy of a Current in a Magnetic Field. — If a circuit carrying a current i is placed in the magnetic field due to a single north pole of strength m, with its positive or north face turned toward the pole, so that it subtends the solid angle Ω as seen from the pole, it will have potential energy equal to $mi\Omega$. This appears at once when we consider that the potential due to the current at the point where the pole is placed is $i\Omega$ (§ 463), and therefore the energy of the pole is $mi\Omega$, and that, since to convert this energy into work the pole and the circuit must be removed out of each other's fields, the same work will be done on the circuit if it is removed while the pole is fixed as will be done on the pole if it is removed while the circuit is fixed.

Now the pole is associated with $4\pi m$ unit tubes of magnetic force, and these radiate from it similarly in all directions, so that $m\Omega$ tubes

of the pole pass through the circuit. Setting $m\Omega = N$, the number of tubes of the pole which pass through the circuit, we may represent the energy of the circuit due to the field of the pole by iN.

Any magnetic field may be constructed by the proper arrangement of magnetic poles, to each one of which the foregoing demonstration applies, so that if N represents the total number of tubes of magnetic force which pass through the circuit, the energy of the circuit in the field will be given by iN.

This energy is positive when the tubes of force of the field pass through the north face of the circuit, or pass through the circuit in a direction in general opposite to that of the tubes of force of the circuit. In that case the potential energy is diminished by moving the circuit so that the number of tubes of force which pass through it diminishes. The circuit will therefore move in the direction in which the most rapid diminution occurs in the number of tubes of force which pass through it.

On the other hand, when the tubes of force of the field pass through the south face of the circuit, or in such a direction that in general their direction coincides with that of the direction of the tubes of force of the circuit, the potential energy of the circuit is negative. The motion of the circuit will therefore be in such a direction that this negative potential energy increases numerically most rapidly, or in such a direction that the most rapid increase occurs in the number of tubes of force which pass through the circuit.

When the tubes of force of the field coincide in direction, in general, with the tubes of force of the circuit through which they pass, they are said to pass through it in the positive direction. The motion of any free circuit in a magnetic field is such as to bring about the result that as many tubes of force of the field as possible pass through it in the positive direction. In a uniform field the circuit will turn until the lines of force of the field pass through it in the positive direction, and there will then be no further motion. In a non-uniform field it will generally turn in a similar manner and also move toward the stronger part of the field.

473. Mutual Energy of Two Currents. — When two circuits carrying currents are in each other's magnetic fields, each possesses energy because of the presence of the other. Since the separation of the circuits to an infinite distance, by which this energy is measured, can be effected by keeping either one at rest and moving the other away, it is evident that the energy of this sort possessed

by each of the circuits must be the same. We may determine the way in which this mutual energy depends upon the currents in the circuits in the following manner:

Let i_1 and i_2 represent the currents in the two circuits 1 and 2; let N_1 represent the number of tubes of force which pass through circuit 1 in consequence of the current i_2 in circuit 2, and N_2, the number of tubes of force which pass through circuit 2 in consequence of the current i_1 in circuit 1. Then, as was shown in the preceding section, the energies of the circuits are given by i_1N_1 and i_2N_2, and these energies are equal. Now the strength of a field due to a current is proportional to the strength of the current, so that we may write $N_1 = M_1i_2$, $N_2 = M_2i_1$, by using M_1 and M_2 as the factors of proportion; and substituting these values in the equation expressing the equality of the energies we obtain $i_1i_2M_1 = i_1i_2M_2$. The factors of proportion M_1 and M_2 are therefore equal to each other and we may use for them the symbol M. The mutual energy of the circuits is then represented by i_1i_2M. The factor M is called the *coefficient of mutual induction*. It evidently represents the number of tubes of force which will pass through either circuit in consequence of the existence of unit current in the other circuit. It depends upon the shape and size of the circuits, their relative positions, and the nature of the surrounding medium.

The potential energy of one of the circuits is positive when the tubes of force of the other circuit pass through it in a direction in general opposite to that in which its own tubes pass through it. If one of the circuits is free to move, it will move so as to diminish its potential energy, and so will move away from the other. If the tubes of force of either circuit pass through the other in the same direction in general as its own, the potential energy is negative. The circuit which is free to move will then move so as to increase this negative potential energy, and so will move toward the other. In this case the tubes of force pass through the circuits in the positive direction. The most general finite motion of one circuit in the field of another will be one in which the circuit will turn so that the tubes of force of the field pass through it in the positive direction and then move into the strongest part of the field or toward the other circuit.

A little consideration will show that these motions are similar to those which the two magnetic shells equivalent to the circuits would execute; and that in general when the circuits move apart, or repel each other, their currents are in opposite senses, while, when they

move together, or attract each other, their currents are in the same sense.

474. Electromagnetic Forces Described by Tensions in Tubes of Force. — The proof which has been given (§ 446) that the electrostatic forces can be explained by ascribing them to tensions existing in the tubes of force of the electrostatic field can be applied to explain forces between magnetic poles, and in general to explain the forces between bodies which are connected with magnetic fields by ascribing them to tensions existing in the tubes of force of the magnetic field. The tension at any point in the field at which the magnetic force is R is along the axis of the tube of force at that point and is equal to $R/8\pi$ for each unit tube. This tension tends to shorten the tubes of force. There is also a pressure equal to $R/8\pi$ acting everywhere perpendicularly to the axis of the tube, and tending to repel contiguous tubes from it. The force on any body in a magnetic field may be explained by supposing it to be the resultant of these tensions and pressures in the tubes of force which are so placed as to act upon the body.

If we can construct a plan of the tubes of force in a field due either to magnets or electric currents we may obtain a notion of the forces which act on the bodies by applying the general rule that the tubes of force tend to shorten in length and to repel one another sideways.

In the case of two parallel circuits in which the currents are in the same sense, the tubes of force of either current are superposed on those of the other in the

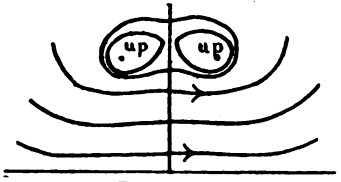


Fig. 267, a.

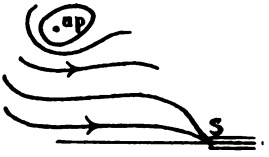


Fig. 267, b.

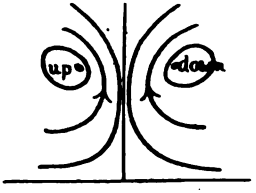


Fig. 267, c.

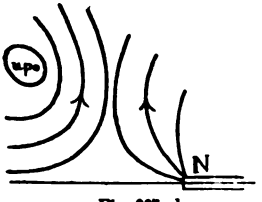


Fig. 267, d.

region between the circuits, so that the plan of the field will be that roughly represented in Fig. 267, a, in which, as in the following cases, only the upper half of the field is shown. In this case, if motion is possible, the shortening of the

tubes of force will draw the circuits together. Similarly, a magnet placed so that its south end points toward the north face of a circuit (Fig. 267, b), will be drawn toward the circuit by the shortening of the tubes of force.

In the case of two parallel circuits in which the currents are in opposite senses, the tubes of force form two separate groups, connected respectively with the two circuits, and lying contiguous to each other along a surface in general transverse to the axis of the circuits (Fig. 267, c). In this case, if motion is possible, the lateral repulsion of the contiguous tubes of the two groups will force the circuits apart. Similarly a magnet so placed that its north end points toward the north face of a circuit (Fig. 267, d) will be forced away from the circuit.

EXAMPLES, XXV

1. *To find the magnetic intensity at the center of a circular current.*

Suppose a unit magnetic pole placed at the center of the circular current, upon which the magnetic field of the current will act. The force upon it, which will measure the magnetic intensity, may be calculated by Laplace's formula.

Let r represent the radius of the circuit carrying the current i. In applying Laplace's formula (§ 461) to this case we notice that the angle θ between any element ds of the circuit and the line joining it with the unit pole is a right angle, so that $\sin\theta = 1$ for each element. The radius r is likewise the same for every term in the summation. We therefore have $R = ci\Sigma ds/r^2$ and $\Sigma ds = 2\pi r$, the circumference of the circle. When the current is measured in electromagnetic units we have further $c = 1$. Hence, finally, $R = 2\pi i/r$.

If the circuit is a circular coil of n turns of wire, the magnetic intensity at the center is $R = 2\pi ni/r$.

2. *To show that Laplace's formula* (§ 461) *leads to Biot and Savart's law.*

Around the point P (Fig. 268), at which the magnetic intensity is to be calculated, draw a semicircle of radius p, which touches the straight current, and stands on a diameter MN parallel with it. We take the constant c of Laplace's formula equal to 1. Consider any element of the current of length ds and at the distance r from the point P. This element will contribute the amount $i\sin\theta\, ds/r^2$ to the total intensity at P. Now $\sin\theta\, ds$ is very approximately the projection a of ds perpendicular to r, and the two lines which are drawn from P to the ends of ds cut out an element b of the circle which is related to a approximately by the proportion $b : a = p : r$. The contribution of the element ds to the intensity may therefore be written as ib/pr. But $b\sin\theta = bp/r = c$, the projection of b on the diameter MN; and using c in the value for the intensity, we have the contribution of the element ds expressed by ic/p^2. The effect of the whole infinite straight current may be represented by the sum of such terms as this; so that the magnetic intensity at P is equal to $\Sigma ic/p^2$, or to $i\Sigma c/p^2$. Now $\Sigma c = 2p$, the diameter of the circle. Hence the magnetic intensity at P equals $2i/p$.

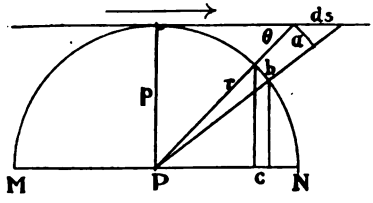


Fig. 268.

This result shows that if the constant c of Laplace's formula equals 1, the constant n of Biot and Savart's formula equals 2.

475. Electromagnetic Induction. — The production of a magnetic field by a current suggested to many observers the possibility of the production of a current by means of a magnetic field. After some years, in which many unsuccessful attempts had been made by himself and others, Faraday, in 1831, succeeded in obtaining the expected action.

In his first experiment Faraday used two concentric spools of wire, one of which could be connected with a voltaic battery. This he called the primary coil. The other, called the secondary coil, was connected with a sensitive galvanometer. With this arrangement Faraday found that, when the current of the battery was thrown into the primary coil, the galvanometer was deflected so as to indicate a current in the secondary coil. This current in the secondary coil was only temporary. While the current in the primary coil was maintained unchanged, the deflection of the galvanometer ceased and no current appeared in the secondary coil. When the circuit of the primary coil was broken, so that the current in it ceased, a deflection in the opposite sense to the one observed before occurred in the galvanometer, indicating a current in the opposite sense in the secondary coil. This current was of course also only temporary. The temporary currents thus set up on making and breaking the primary circuit were called by Faraday *induced currents*, and they were said to be produced by *electromagnetic induction*. By observations of the deflections of the galvanometer Faraday proved that the sense in which the induced current circulates in the secondary coil is opposite to that of the current in the primary coil when the primary circuit is made, and is the same as that of the current in the primary coil when the primary circuit is broken. Fig. 269, a represents the case in which the inner circuit, the primary, is made.

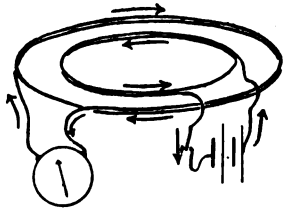


Fig. 269, a.

Reasoning that the setting up of a current in the primary coil is, in effect, bringing the primary coil to its position near the secondary from an infinite distance, Faraday next tried the experiment of moving the primary coil, while the current was established in it, up to the secondary coil. He found then that the galvanometer was again deflected, showing a current in the secondary coil, in the same sense as that of the current produced when the

current was thrown into the primary coil in the first experiment. When the primary coil was rapidly removed from the neighborhood of the secondary coil, a current was produced in the secondary coil in the opposite sense. Fig. 269, b represents the case in which the upper circuit, the primary, is moved toward the secondary.

Faraday next substituted a magnet for the primary coil, and found that when it was moved up to the secondary in the proper way, and especially when it was thrust into the secondary for half its length, an induced current was developed, and that an induced current in the opposite sense was developed when the magnet was removed. The direction of the induced currents produced by the movement of a magnet may be readily obtained from the rule already given, by remembering the relation between the lines of force of a magnet and the direction of the current in the solenoid which is equivalent to the magnet. When the north pole of the magnet is thrust into the coil, the sense of the current in the coil, looked at from the side on which the magnet is, is counterclockwise. Fig. 269, c represents this case.

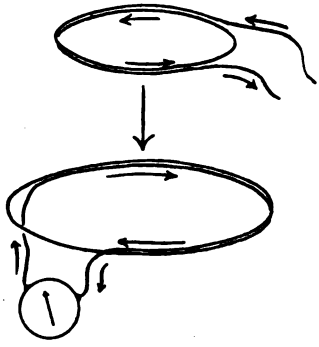


Fig. 269, b.

Faraday considered the induced current to result from the change of the magnetic field around the circuit in which it is produced. He looked on a magnet as a body which carries with it a set of lines of magnetic force, and he described the production of the induced current by the movement of a magnet as occurring whenever the circuit in which it arises cuts through lines of force. It is perhaps easier for us to consider the induced current as produced by an alteration in the magnetic field enclosed by the circuit. The induced current is produced only while the field is changing, and ceases when the field becomes constant.

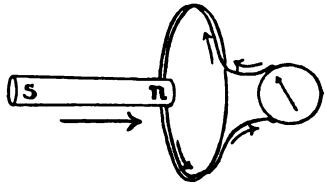


Fig. 269, c.

The principal phenomena of induced currents were discovered by Joseph Henry about the same time that they were discovered by Faraday and independently of him.

476. Self-induction. One day after a lecture Faraday had his attention called by a gentleman in the audience to a fact which he had observed, that when a circuit is broken in which a coil of wire is contained, a bright spark appears at the break. On investigating this phenomenon, Faraday found that the spark is brighter when the wire in the circuit is in the form of a coil than when the same wire is in the circuit but is not coiled up; and reflecting on this, he perceived that the spark is due to an induced current set up in the circuit itself by the change in its own current. The action to which this effect is due is called self-induction. Self-induction was also discovered and studied by Henry. If we apply the rules for the production of the induced current to a single circuit, considering it to act as both primary and secondary circuit at once, we see that when a current is thrown into the circuit, an induced current will be developed in the opposite sense. The effect of this current is to retard the full development of the primary current, so that the current will be established more slowly in a circuit which is coiled up so that its parts can act inductively on one another, than it will be in the same circuit if it is not coiled. When the circuit is broken, the departure of the current produces an induced current in the same sense as that of the original current, so that a momentary current of greater strength is developed. It is to this current that the spark is due which is observed when a circuit is broken. It is commonly called the *extra current*, and sometimes the current of self-induction.

477. Lenz's Law. — A general rule, by which the direction of the induced current in a circuit may be determined, was given by Lenz, and is known as Lenz's law. If we examine the direction of the induced current produced by bringing a primary current toward the secondary circuit, we notice that it is in the opposite direction to the current which would attract the primary current. The induced current in this case therefore tends to repel the primary current, or to oppose the work which is being done on the primary circuit. In all other cases the same general rule holds, and we may state Lenz's law by saying, that the sense of the induced current is always such that the force exerted by the current opposes the action by which it is produced.

An interesting experiment which exhibits Lenz's law was performed by Arago, several years before the induced current was discovered. He arranged a copper disk so that it could be rotated rapidly around a vertical axis, and suspended above it a long magnet. When the disk was rotated the magnet was deflected

in the sense of the rotation, and when suspended so as to be capable of free rotation around the axis of the disk, it rotated in the same sense as that in which the disk was rotated. Arago could not explain this experiment, and it remained unaccounted for until Faraday discovered induction. It was then explained by Faraday as a consequence of the induction of currents in the copper disk by its movement in the magnetic field of the magnet. According to Lenz's law, the currents thus induced are all in such a sense as to oppose the motion producing them, and hence there arises a force between the disk and the poles of the magnet tending to prevent the rotation of the disk, or what amounts to the same thing, to set up a rotation of the magnet in the sense of the rotation of the disk. That currents of the sort assumed in this explanation really exist in the disk can be shown by rotating a disk in a strong magnetic field, and touching two parts of it with the terminals of wires leading to a galvanometer. With this arrangement a continuous current will pass through the galvanometer so long as the disk is rotated. The arrangement is therefore a machine for the production of the electric current by motions in a magnetic field. The first model of such a machine was constructed by Faraday.

478. The Dynamo Machine. — The extensive use of the electric current in our modern life originated in this discovery of Faraday's, by which it is made possible to use mechanical energy directly, and on a large scale, for the production of the current. The so-called *dynamo machine*, used for this purpose, consists generally of a large fixed electromagnet, which furnishes a strong magnetic field, and of a rotating portion, called the armature, made of a properly shaped mass of iron, over which wires are wound in such a manner that as the armature is rapidly rotated by an engine in the magnetic field of the field magnets, a current is developed in them. By means of sliding contacts between these wires and the external circuit, the current thus developed is led off from the machine. When the current is led into a similar machine the interaction between the magnetic fields of its various parts sets up and maintains a rotation of its armature. The machine thus used is called a *motor*.

Other forms of the dynamo machine, which produce an alternating instead of a direct current, and other motors which can be operated by the alternating currents, have been devised and are now very extensively used.

479. Induction Coil. — The *induction coil*, called also the Runmkorff coil, or the inductorium, is an instrument used to produce induced currents with high electromotive forces. It consists of a cylindrical primary coil of coarse wire wound around a central core of iron, and a secondary coil, outside the primary and concentric with it, containing very many turns of fine wire. An

automatic circuit breaker is employed, by which the current from a battery is rapidly made and broken in the primary circuit. By the addition of a condenser to the primary circuit, which is charged when the circuit is made and broken, the development of the primary current is delayed and its annihilation is hastened, so that the electromotive force of the induced current in the secondary coil, set up on making the primary circuit, is depressed, and that of the current set up on breaking the primary circuit is heightened. The consequence is that the induced current formed at the breaking of the primary is able to leap over larger gaps than the other, and so, when the current from the coil is sent through a nonconductor, like a rarefied gas, it acts almost as if it were passing only in one direction.

Electrochemistry

480. Electrolysis. — The decomposition of chemical compounds by the electric current, and the deposition of the constituents of the compound on the solid terminals by which the current was introduced into it, were used by Davy and others in chemical analysis, and led to the discovery of several elements and to new views of the nature of chemical composition. No accurate knowledge of the chemical relations of the current was obtained, however, until 1834, in which year Faraday established the science of electrochemistry by his discovery of the laws which govern the relations between the electric current and the chemical effects produced by it.

At the outset, starting with the general knowledge of the chemical action of the current which has already been described (§ 456), Faraday introduced a new set of terms to describe the action under investigation. These terms or names did not commit him to any theory of the action. They are simply descriptive, and were so well fitted to their purpose that they have always been employed since they were introduced. In this nomenclature the substance which is decomposed by the current is called the electrolyte. When decomposed it is said to be electrolyzed, and the process of decomposition is called *electrolysis*. The two terminals by which the current is introduced into the electrolyte are called the electrodes, the positive one, or the one at higher potential, being called the anode, the other the cathode. The products of electrolysis are called ions, the one which appears at the anode being called the anion, the other the cation.

481. Faraday's First Law of Electrolysis. — Faraday devoted special attention to the electrolysis of water. He had no means of measuring the strength of the current which he used, but by altering various features of the circuit, in such a way as not to change the current appreciably, for instance by changing the size of the electrodes and by electrolyzing different solutions, he convinced himself that the amount of electrolysis is independent of everything connected with the circuit except the quantity of electricity which passes. As we should now put it, the amount of electrolysis is proportional to the quantity, and the rate at which electrolysis takes place is proportional to the current strength. Faraday generalized this conclusion into the *first law of electrolysis*, which we may state by saying that the amount by weight of an ion evolved at an electrode during the passage of a current for unit time is proportional to the current strength. After galvanometers were constructed by which an independent measure of current strength was obtained, this law was investigated with the utmost care and completely verified.

482. Faraday's Second Law of Electrolysis. — Faraday also investigated the electrolysis of different compounds, yielding different chemical elements as ions, by the passage of the same current. He discovered by these experiments a most important relation. To state it, it will be necessary to say a word about the relations of the chemical elements to one another.

From the relations of the weights of chemical compounds to one another chemists have agreed that the chemical elements exist in the form of minute atoms, which combine with one another to form chemical compounds, and that the atom of any particular element has a definite weight, which is characteristic of that element. The weight of the hydrogen atom is usually taken as the standard or unit weight, and the weight of the atom of another element, expressed in terms of that standard weight, is called its *atomic weight*.

When chemical compounds are examined it is found that they generally consist of molecules containing a few atoms of the elements which form the compound. In many instances an element may be removed from such a compound, and another substituted for it. When this is done, it is often found that the place vacated by one number of atoms of one element is filled by a different number of atoms of the other element. By such experiments it may be determined how many atoms of one element are equivalent

in their combining power to a single atom of another element. If we conceive such experiments to be carried out by replacing hydrogen atoms by a single atom of another element, the number of hydrogen atoms displaced by the one atom of the other element is called the *valency* of that element. Sometimes a small group of atoms of different sorts, forming what is called a radical, will replace a number of hydrogen atoms in a compound. In such a case the radical has a valency equal to the number of hydrogen atoms which it replaces.

If we suppose an element having one valency to replace in a compound another having another valency, it is plain that the weights of the two elements which exchange places will be proportional to their atomic weights divided by their valencies. The ratio of the atomic weight to the valency is called the *chemical equivalent* of an element.

We are now in a position to state the *second law of electrolysis*, by saying that the weights of the ions produced at the various electrodes, when a current passes through different electrolytes, are proportional to the chemical equivalents of the ions.

To illustrate this law let us suppose that the same current is used to electrolyze water, sulphate of copper, and chloride of silver. The different elements concerned, which we need to consider, are arranged in the following table, with their respective atomic weights, valencies, and chemical equivalents:

Elements	*Atomic Weight*	*Valency*	*Chemical Equivalent*
Hydrogen	1	1	1
Oxygen	16	2	8
Copper	63	2	31.5
Silver	108	1	108
Chloride	35.5	1	35.5

When the same current is sent through these compounds for a while, and the products of electrolysis collected and weighed, it will be found that, for every gram of hydrogen, there are obtained 8 grams of oxygen, 31.5 grams of copper, 108 grams of silver, and 35.5 grams of chloride.

483. Ionic Charge. — A conclusion can be drawn from the facts embodied in the second law of electrolysis which gives us an insight into the nature of distribution of electricity. When we consider the exact proportion between the current-strength and the weights of the ions which are evolved at the electrodes, we are led to consider the passage of the current from the electrolyte to the electrode as effected by the transfer of electric charges to the electrode by the

ions which are developed on them. An ion is made up of units, either atoms or radicals, which are alike in every respect. These units are called *elementary ions*, or often simply *ions*. Each elementary ion which has formed a part of a molecule of the compound brings to the electrode a certain number of units of valency. The second law shows that, for a given current, the ions of different compounds bring to their respective electrodes the same number of units of valency. If, therefore, we suppose electricity to be divided into equal portions, which we may call *ionic charges*, and if we suppose that one such ionic charge is associated with each unit of valency, and that the charges of the ion are surrendered to the electrode when the ion is developed on it, we then can account for both the laws of electrolysis. It is plain that this conception is inconsistent with any conception of electricity as being of the nature of a continuous fluid.

484. Grotthus's Theory of Electrolysis. — Within a few years after the discovery of electrolysis, a theory was advanced by Grotthus to explain it, from which the more modern theories were developed. Grotthus supposed that each molecule of the electrolyte is made up of two equally and oppositely charged parts, which become the ions when the molecule is broken up. When a difference of potential exists between the electrodes, so that there is an electric force in the electrolyte, these molecules arrange themselves in chains or rows, with their positive charges all pointing in one direction, toward the negatively charged electrode, or in the direction of the current. The molecules of the rows are then broken into their ions by the electric force, and the positive ions move toward the cathode, the negative toward the anode, thus setting free an ion at each end of the row, and bringing the remaining positive and negative ions together so that they form new molecules. These new molecules thus formed then arrange themselves as before under the electric force and the process is repeated. This theory or description accounts for the evolution of the ions at the electrodes and not in the body of the electrolyte, and it indicates a relation between the amount of electrolysis and the strength of the current. When Faraday's laws were discovered, he found that they also could be explained by Grotthus's theory, although Faraday himself preferred another one, which was not, however, essentially different.

485. Williamson-Clausius's Theory of Electrolysis. — When electrolysis was studied more carefully, certain phenomena were

observed which could not easily be reconciled with this theory. In particular, it was found that no matter how small the difference of potential is between the two electrodes, a current will pass between them through the electrolyte, and electrolysis will occur. To account for this and for certain other peculiarities observed, as, for example, the fact that the conductivity of an electrolyte increases with the temperature, Williamson in England and Clausius in Germany modified the theory of Grotthus. In their form of the theory use was made of the hypothesis that the molecules of all bodies are in active motion. This being admitted, it was assumed that occasionally, when two molecules of the electrolyte encounter each other, they are so affected by their mutual electric forces, or by the shock of impact, as to be broken into their constituent ions. These ions exist in the electrolyte for a time as free ions, and during that time move toward the one electrode or the other, according as they are positive or negative. This description of electrolysis is no doubt a considerable advance on the one which we first examined.

486. Dissociation Theory of Electrolysis. — A further modification of the theory was made in consequence of a discovery of Kohlrausch. By determining the currents in different electrolytes under similar conditions, Kohlrausch showed that they can be accounted for by supposing that the ions of each particular sort, when urged by a given electric force, move through the electrolyte with a velocity which is peculiar to that sort of ion, and which is the same for that ion from whatever combination it is evolved. Thus, as we may say, the conductivity of an electrolyte, that is, its power of conveying current under standard conditions, is the sum of two numbers characteristic of its ions.

Another experimental result was also influential in leading to a modification of Clausius's theory. It was found that, during the process of electrolysis, those parts of the electrolyte near the electrode contain more than the average amounts of the two ions. This movement of the ions toward their respective electrodes was investigated by Hittorf. The relative rates of movement for the different ions were found by Kohlrausch to agree with those which he had determined from his observations of conductivity.

The requisite modification of the theory of electrolysis, to account for these facts, was made by Arrhenius. The theory which he developed is called the *dissociation theory*. Arrhenius supposes that

when a salt or any other substance is dissolved so as to form an electrolyte, a certain proportion of its molecules separate into their ions, so that a solution, whether acted on by electric forces or not, contains a large number of such free or dissociated ions. With each of these ions there are associated electric charges of the proper sign, and proportional in quantity to the valency of the ion. The effect of introducing an electric force in the electrolyte is to cause a drift of these free ions toward the electrodes. When they reach the electrodes, they give up their charges to them and are evolved as the products of electrolysis (Fig. 270). As the free ions are successively removed in this way from the electrolyte, other molecules of the solute are dissociated, so that the supply of free ions is kept up. The fundamental assumption of this theory, that a considerable proportion of the molecules of the solute are dissociated in the solution into two parts, is borne out by the effect which such solutes have in lowering the freezing point and in raising the boiling point of their solutions (§ 239).

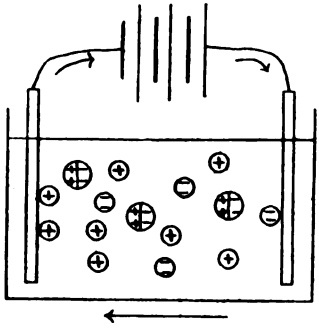


Fig. 270.

487. The Voltameter. — Faraday suggested the use of the electrolysis of water as a means of measuring the quantity of electricity which passes through a circuit. The instrument which he proposed he called the *voltameter*.

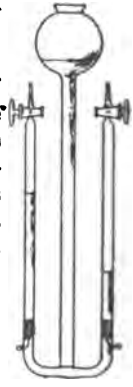


Fig. 271.

In one of its many forms, the voltameter (Fig. 271) consists of two graduated tubes, which stand vertical and are connected at the bottom by a short horizontal tube. To this connecting tube is joined another vertical tube in which the superfluous water is collected which is forced out as electrolysis proceeds. Platinum plates are introduced into the two graduated tubes and joined to wires carried through their walls. The apparatus, when ready for use, is filled with acidulated water. When the current to be measured is passed between the electrodes, electrolysis begins, and the two ions, oxygen and hydrogen, begin to collect in the upper parts of the tubes. The quantity of electricity which passes while the current is allowed to flow is then determined by the amount of oxygen or hydrogen which is obtained, and if the current is kept constant while it is flowing, the amount of either gas evolved in unit time measures the strength of the current. In the use of this instrument by Faraday, the current indicated by it was measured in terms of an arbitrary unit, namely, that current which will evolve a determined quantity of gas in a fixed time. In the practical use of

the voltameter we still often employ such a unit, considering, for example, that current to be unit current which will evolve a cubic centimeter of hydrogen in one minute. As we now, however, know the absolute value of the current which will produce this result, we commonly express the strength of a current in absolute units, even when we have used the voltameter to measure it.

It is obvious that any other electrolytic process may be used as a means of measuring currents. In refined work it has been found best to use the electrolysis of silver from a solution of nitrate of silver. The relation of the absolute value of the current to the amount of silver which it will deposit from such a solution has been very carefully determined. It is agreed that the ampere will deposit 1.118 milligrams of silver in one second.

488. The Voltaic Cell. — Now that we have studied the process of electrolysis, we are in position to examine the action of the voltaic cell. As now becomes apparent, the conductor of the second class which is interposed between the two plates or poles of the cell is an electrolyte. When the poles are joined by a conductor other than this electrolyte, an unbalanced difference of potential exists in the circuit, and a current passes through the electrolyte. This current in the electrolyte (Fig. 272) is from the electropositive element of the cell to the electronegative element, that is, in the typical case already described, from the zinc to the copper, and as it flows, the electrolyte between these elements is broken into its ions. The anion is evolved on the electropositive element, the cation on the electronegative element. Now it always is the case, in any voltaic cell, that the anions which reach the electropositive element combine with its atoms to form molecules, which are dissolved in the electrolyte, so that the electropositive element, which is the negative pole of the battery, wastes away. The cation, which is evolved at the other element, will sometimes interact with it or with the solvent and sometimes not. If it does not, it appears as the product of electrolysis at the positive pole. If it does interact with that element or with the solvent, some other substance is evolved by that action as a product of electrolysis.

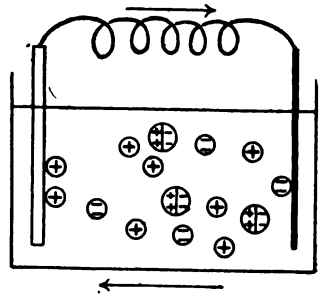


Fig. 272.

To illustrate this general description, we may consider the simple voltaic cell formed of plates of copper and zinc, immersed in a solution of sulphuric acid.

When sulphuric acid is dissolved, it dissociates into two positive hydrogen ions and a negative ion, called the sulphion, containing one atom of sulphur and four of oxygen. Each of the hydrogen ions is univalent, and carries one positive ionic charge. The sulphion is bivalent and carries two negative ionic charges. When the circuit is joined, a difference of potential exists between the zinc and the copper plates, the potential of the zinc being the higher. The hydrogen ions move toward the copper plate, and are evolved at it without acting on it chemically. The sulphions move toward the zinc plate, and when they reach it, combine each with an atom of zinc, so as to form sulphate of zinc, which dissolves in the solution. Nothing appears at the zinc plate as a product of electrolysis, and the plate gradually wastes away.

489. Theory of the Voltaic Cell. — In giving this description of the actions in the cell, the order of events has been left as uncertain as possible. To make a definite statement about it would involve deciding between two rival theories of the action of the cell, known respectively as the contact theory and the chemical theory. In the *contact theory*, which was proposed by Volta, and which has been highly developed by Lord Kelvin, the action of the cell is ascribed to the unbalanced potential difference in the circuit, arising from the contacts of the different parts of the circuit, and the chemical action, by which the zinc is decomposed, results as a consequence of electrolysis set up by this potential difference. In the *chemical theory*, held by Faraday, the potential difference in the circuit is ascribed to a tendency to chemical action between the zinc and the sulphions, by which a condition of strain is set up around the zinc, and indeed throughout the cell. When the circuit is joined, this strain can be relieved by the flow of electricity around the circuit. Experiment has not yet decided between these two theories.

490. Polarization. — When a cell of the sort just described is set in operation, the current developed by it gradually diminishes in strength, until it becomes much weaker than it was at first. This result is ascribed to what we call the *polarization* of the cell. The hydrogen evolved at the copper plate collects on it in a very thin layer, and so, in effect, partially replaces the copper plate by a plate of hydrogen. Not only is this layer of hydrogen a very poor conductor, but also the difference of potential between it and zinc is less than that between copper and zinc. For both these reasons the current of the cell is weakened. By using platinum instead of copper, by roughening the surface of the platinum so that the hydrogen evolved on it more readily forms bubbles and escapes, and by shaking the plate, much of the hydrogen can be removed

mechanically, and a cell obtained in which the polarization is not great.

Polarization may be avoided also by using other combinations of materials, so selected that the ion which appears at the positive pole is of the same sort as the pole itself. Thus in the Daniell's cell, the ordinary cell used in telegraphy, the plates are of zinc and copper, and two liquids are used, a solution of sulphate of zinc around the zinc, and a solution of sulphate of copper around the copper. These liquids are separated from each other by a porous jar, or by the difference in their specific gravities. When the current passes through this cell, it evolves sulphions at the zinc, so that the zinc is reduced, and copper ions on the copper, so that the copper plate merely becomes thicker, but without the character of its surface changing. At the surface where the two solutions meet, zinc ions proceeding toward the copper plate meet with sulphions proceeding in the opposite direction, and combine to form sulphate of zinc, so that an additional effect of the action is to diminish the quantity of sulphate of copper and increase the quantity of sulphate of zinc. In order to prepare such a cell so that it will furnish current for a long time without charging, crystals of sulphate of copper are placed in the solution of sulphate of copper, so that the strength of that solution may be kept up.

Other forms of cell are used in which polarization is avoided by surrounding the positive pole with some substance which acts chemically upon the ions evolved, and removes them from the pole. The Leclanché cell and the dry battery, used so much for ringing electric bells and for other work of that kind, are examples of such cells. In them the depolarizing material works slowly, so that the current cannot be run for a long time without polarizing. But when the cell is used only for a short time, it furnishes a strong current, and it is depolarized by the time that it is again needed.

The storage cell, developed from a cell constructed by Planté, contains two lead plates, coated with oxide of lead and immersed in sulphuric acid. When a current is passed through this cell, a principal part of the action consists in a reduction of the oxide on one plate to spongy lead, and the conversion of that on the other plate into a higher oxide. When the terminals of the cell are joined, after it has thus been charged, the current which it delivers is accompanied by a return of the surfaces of the plates to their initial condition.

THERMOELECTRICITY

491. The Thermoelectric Current. — In our study of Volta's series it was stated that no current will be set up in a circuit formed exclusively of conductors of the first class, or that, in other words, such a circuit will be in electric equilibrium. It was discovered in 1822 by Seebeck that this equilibrium is disturbed when one of the junctions of the conductors composing this circuit is heated. In that case a continuous current flows in the circuit, so long as the difference of temperature between the heated junction and the others continues. The current thus formed is called the *thermoelectric current.* The strength of the thermoelectric current depends on the difference of temperature between the heated junction and the others. It also depends upon the nature of the conductors at the heated junction. A pair of such conductors, arranged for the production of the thermoelectric current, is called a *thermocouple* or a *thermoelement.*

The conductors are usually metals, and for each pair of metals the current which is developed may be accounted for by supposing that a difference of potential exists at the point of contact between the metals, and that this difference of potential changes with the temperature. This difference of potential is not that discovered by Volta, but very much less than that. The rate at which this difference of potential changes with the temperature is called the *thermoelectric power* of the couple. That conductor of the couple from which the thermoelectric current flows to the other, across the heated junction, is said to be thermoelectrically positive to the other conductor. Thus, for example, when a couple is formed of antimony and bismuth (Fig. 273), it is found by experiment that the current flows from the antimony to the bismuth across the heated junction; the antimony is therefore thermoelectrically positive to the bismuth. Experiment shows that all the metals can be arranged in a series, such that each successive member of the series is thermoelectrically positive to those which follow it, and thermoelectrically negative to those which precede it.

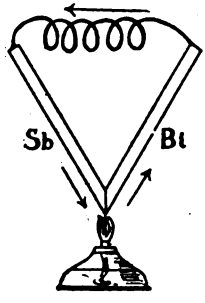


Fig. 273.

The current produced in an ordinary circuit by a single thermoelectric element is very slight, but a number of such elements may be joined in succession, so that their differences of potential are

added to one another, and the current produced by them when their alternate junctions are heated may be very considerable. Such an arrangement of elements, called a *thermopile*, is a very sensitive instrument for the detection of radiant heat.

492. The Peltier Effect. — It was discovered by Peltier in 1834 that when a current from a battery or any other source is passed through the junction of two metals in the sense of the current which would be set up by heating that junction, the junction becomes cooler. If the current is passed in the opposite sense, the junction is heated. This effect is known as the *Peltier effect.*

General Laws of the Electric Current

493. Identity of Electricity from Different Sources. — The course of Volta's thought during the investigations which resulted in his invention of the voltaic battery naturally led him to believe that the effects produced by the battery were electric effects, and that the battery was essentially equivalent to a set of Leyden jars from which a continuous discharge was passing. This supposed continuous discharge in the circuit of the battery was called the electric current.

The reasons which Volta had for this view were, first, that a difference of potential could be shown to exist between the two poles of the battery, similar in kind to that existing between the two coatings of a Leyden jar, though generally very much less than that in degree, and second, that both the current and the discharge from the Leyden jar heated the conductors through which they passed.

It was felt by Faraday that additional proof of this view was desirable. He accordingly investigated the discharge of the Leyden jar, in order to determine whether or not all the effects which are produced by the current are also produced by it. He found that, by proper arrangements, the discharge of the jar could be made to produce chemical decomposition. The amount of chemical action was so slight that it could only be detected by him by a special artifice. He impregnated a piece of blotting paper with a solution of starch and of iodide of potassium, and sent the discharge through it between two terminal wires whose ends touched the paper. When the discharge had passed, a small blue spot appeared around the end of one wire, due to the action upon the starch of the iodine released by the chemical action of the discharge. The

evolution of the iodine occurred at the terminal at which it was expected to occur from what was known about the action on the same compound of the electric current from a battery, and convinced Faraday of the essential similarity of the discharge and the current. With our modern electric machines a continuous chemical decomposition can be maintained. The amount of this decomposition, even with the most powerful machines, is extremely slight in comparison with that produced by a battery.

Faraday also showed, by discharging Leyden jars through a circuit containing a galvanometer, that the discharge affected the magnetic needle, or produced a magnetic field. The same thing was shown by Joseph Henry, who magnetized sewing needles by placing them in the axis of a solenoid and sending the discharge through it.

From these experiments, and many others in which the currents obtained from various sources were compared, Faraday concluded that electricity from all these different sources was identical in kind.

494. Discharge of a Condenser. — When a condenser, or in fact any insulated conductor, is charged, we conceive it to contain or carry on its surface a quantity of electricity Q. Its potential V is different from the potential V_0 of the conductors which surround it. An equal and opposite charge $-Q$ resides on these conductors. In these circumstances the electric energy of the system is $Q(V - V_0)/2$ (Examples, XXIII, 13). This energy is transformed into other forms of energy, principally into heat, when the condenser is discharged.

It is not material to the study of the electric current that we should form a correct conception of all that takes place during the discharge. On the one-fluid theory of Franklin, a stream of electricity moves from the positively charged body to the negatively charged one until electric equilibrium is attained. On the two-fluid theory of Dufay, equal quantities of positive and negative electricities move in opposite directions along the conductor through which the discharge takes place, until a neutral state is attained, throughout the system.

The discharge is not generally a single and simple flow of electricity from one conductor to another. Without touching the question at this point of the part taken by the dielectric in the discharge, it should be mentioned that in most cases the discharge of a condenser is oscillatory in character, consisting of a series of discharges, which occur at equal intervals of time in opposite senses, and diminish in intensity until they become imperceptible. This feature of the discharge was first recognized by Joseph Henry, in connection with observations

which he made on the magnetization of sewing needles placed in the axis of a coil of wire through which the discharge was passed. Henry found that the needles which he used in successive trials, in which the discharge was sent in a similar manner through the coil, were magnetized sometimes in one sense, sometimes in the other, as if the magnetizing current in the coil had traversed it in opposite directions. Henry perceived that this result would follow from an oscillatory discharge of the character which has been described. A steel magnet placed in a magnetic field in which the lines of force are oppositely directed to the lines of induction of the magnet will retain its original polarity, owing to what is called its coercive force, unless the strength of the field exceeds a certain value. This being so, we may explain Henry's experiment by supposing that the needle is magnetized by the first rush of the discharge, perhaps reversed by the return rush, and so on, until the field set up by one of the waves is too feeble to reverse the polarity which the needle received from the preceding one. The needle will then be found to have the polarity which is given it by the last wave which sets up a magnetic field strong enough to reverse the effect of the field which immediately preceded it.

The oscillatory character of the discharge in most cases was afterwards demonstrated theoretically by Lord Kelvin, and its intermittent character was exhibited by Feddersen, who examined the electric spark in a rapidly revolving mirror and found it to consist of a number of separate sparks, occurring at the ends of equal intervals of time, and rapidly diminishing in intensity.

This peculiar feature of the discharge is of great theoretical and practical importance, but is not important in the use which we are to make of the analogy between the discharge and the current. We may for the present think of the discharge as the passage of electricity from a place of higher potential to a place of lower potential, without scrutinizing more carefully the details of the process.

495. The Electric Current. — From the similarity which exists between the effects of the discharge and the effects of the current, we are led to conceive of the current as a passage of electricity along the conductor in which and around which these effects are exhibited. If the current is constant and if Q units of electricity pass through any cross section of the conductor in the time t, the quantity $I = Q/t$ will pass through the same cross section in unit time. This amount of electricity, the amount which passes through a cross section of the conductor in unit time, is called the *current strength*, or often simply the *current* in the conductor. The current measured in this way is said to be measured in electrostatic units.

The *electrostatic unit* of current is the current of such strength that while the current is constant, one electrostatic unit (§ 424) of electricity passes through a cross section of the conductor in one second.

By testing the various parts of a circuit with a magnet and otherwise, Faraday showed that the strength of the current in a circuit is the same in all parts of it.

496. Electromotive Force. — In many cases, tests made upon the two ends of a homogeneous part of a circuit, such as a wire joining the poles of a battery, show that these ends are at different potentials. If the charge Q traverses this conductor in the time t, the energy expended by its passage from the higher potential V to the lower potential V_0 is $Q\,(V - V_0)$, and the energy expended in unit time is $(Q/t)\,(V - V_0) = I\,(V - V_0)$. This energy appears in such a homogeneous conductor in the form of heat, and can be measured as such. Using W to represent the amount of energy expended in unit time, we have

$$W = I\,(V - V_0). \tag{141}$$

If we measure the difference of potential and the heat developed in the conductor, we may measure the current by the help of this formula, and avoid the necessity of using the electrostatic unit of quantity directly.

In a complete circuit in which a current exists, energy is being expended at a certain rate. This is conceived to be due to a property of the battery or, more generally, of the circuit, which is called its *electromotive force*. The electromotive force is so far analogous to the difference of potential described in the preceding paragraph that it conforms to an equation of the same form. We define the electromotive force as the measure of the expenditure of energy in a circuit. Representing it by E, and the energy expended in the circuit in unit time, or the rate of expenditure of energy, by W, we have

$$W = IE, \tag{142}$$

as the relation which measures the electromotive force of the circuit.

The electrostatic unit of electromotive force is the electromotive force in a circuit in which the electrostatic unit of current will expend one erg in one second. A difference of potential may be measured in the same unit by a similar definition.

497. Resistance. — When experiments were tried with voltaic batteries of different sorts, it was soon perceived that the strengths of current obtained in different cases depended on the character and arrangement of the voltaic cells employed, and, for the same set of cells, upon the character and arrangement of the external circuit. These currents were at first roughly classified as intensity currents and quantity currents. The intensity current was pro-

duced by such an arrangement as the voltaic pile, or a voltaic battery containing a large number of cells in succession, or *in series*, so that the difference of potential between the two terminals, before the circuit was made, was high. It was found that such an arrangement would develop in a long wire a current which did not differ very much from that which it would develop in a shorter one. The quantity current was developed by a single large cell, or by a combination of cells with their like poles joined, or *in parallel*, so as to be equivalent to a single large one. Such a battery proved, by the amount of electrolytic action which it would perform in a short circuit, that it would send out a large quantity of electricity, but it was not capable of developing a considerable current in a long circuit.

The battery in any one of these cases presumably furnishes a constant electromotive force, and the question that is raised by these experiments is therefore that of the dependence of the current strength in a circuit in which the electromotive force is constant upon the materials constituting the circuit and their arrangement in it. The law which governs this relation was discovered in 1826 by Ohm. The currents upon which he experimented were developed in their circuits by a thermopile kept at a constant temperature, which served to maintain a constant electromotive force in whatever circuit it was placed. By their actions on a magnet he measured the strengths of the currents developed in his circuits as their lengths and other characteristics were altered, and concluded, in effect, that the current strength is directly proportional to the electromotive force in the circuit, and that the factor of proportion, called the *conductance*, is independent of the electromotive force, and depends upon the length and cross section of the circuit and the materials which compose it. In the following year, in the absence of exact methods of measurement by which this law could be tested readily, Ohm presented it as the result of speculation on the nature of the electric current. It was carefully studied and verified by Pouillet in 1837, by the help of the tangent galvanometer, and has been abundantly confirmed since that time by the most exact observations.

If we represent the conductance by C we may write the relation proposed by Ohm in the form $I = CE$. The reciprocal of C is commonly used in place of C. It is called the *resistance* and denoted by the symbol R. With this change the formula becomes

$$I = \frac{E}{R}, \tag{143}$$

and embodies what is known as *Ohm's law*, — that the current strength in a circuit equals the electromotive force divided by the resistance.

Similarly, in a limited part of a circuit, we have

$$I = \frac{V_1 - V_0}{R}; \tag{144}$$

the current strength equals the difference of potential between the ends of the conductor divided by its resistance.

Since we can measure the current and the electromotive force in electrostatic units, Ohm's law enables us to measure the resistance. We may further define the electrostatic unit of resistance as the resistance of a circuit in which the electrostatic unit of electromotive force maintains an electrostatic unit of current.

In connection with the subject of resistance it may be stated that the resistance of any cylindrical or columnar conductor, like a metallic wire or rod, or the column of acidulated water in a trough or battery, is found to be proportional to the length of the conductor and inversely proportional to its cross section. It is also proportional to the *specific resistance* or *resistivity* of the substance, that is, the resistance of a cylinder of the same substance of unit length and unit cross section.

498. The Electromagnetic System of Units. — By making use of the relations of current to electromotive force and to resistance by which the latter quantities have been defined, but starting with a different concept of current, and therefore with a different unit of current, we may define other units of electromotive force and of resistance. The concept of current on which this set of units is based is that it is an action in the conductor or circuit which sets up the magnetic field about it. The unit current strength is then defined as the current strength which will set up a magnetic field around a circuit similar to that of a magnetic shell of unit strength, whose edge coincides with the circuit.

We use small letters to represent quantities measured in this system of units.

Defining electromotive force as before as the ratio of the rate at which energy is expended in the circuit to the current strength in it, so that the rate of expenditure of energy w, the current strength i, and the electromotive force e conform to the formula

$$w = ie, \tag{145}$$

we define the unit electromotive force in this system as the electro-

motive force which is associated with unit rate of expenditure of energy in a circuit in which there is unit current strength.

Also defining resistance by Ohm's law as the ratio of the electromotive force to the current strength, so that the current strength i, the electromotive force e, and the resistance r conform to the formula

$$i = \frac{e}{r}, \tag{146}$$

we define the unit resistance in this system as the resistance of a circuit in which unit electromotive force will give rise to unit current strength.

With these concepts thus defined, all other concepts which will be encountered in the development of electric science may be consistently defined and measured. The system thus developed is called the *electromagnetic system* of units.

499. Units of Current Strength. — Both the electrostatic and electromagnetic systems are systems of *absolute* units; that is, the units defined in the system ultimately involve in their definitions, besides the physical laws used in defining them, only the fundamental arbitrary units of mass, length, and time (§ 4).

To perceive this in the case of the electrostatic unit of current we notice that the definition of that unit (§ 495) involves electric charge and time; and turning to the definition of electric charge (§ 424), we see that it involves only force and distance. Now force, a purely mechanical concept, is measured in units of mass, length, and time, so that the definition of unit current in the electrostatic system involves those units only.

To present this argument more formally we shall proceed to construct the *dimensional formula* of current strength in the electrostatic system. When electric charge is measured in electrostatic units (§ 424) the product of two charges divided by the square of the distance between them equals the force with which they act on each other, so that the dimensions of the product of two charges are those of a force multiplied by those of the square of a distance. The dimensional formula of force is MLT^{-2}, so that the dimensional formula of the product of two charges is ML^3T^{-2}; and the dimensional formula of electric charge is therefore $M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$. Since current strength is measured by the ratio of electric charge to time (§ 495), the dimensional formula of electric current is finally $M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}$. This formula shows that the measurement of electric current involves no arbitrarily chosen units except those of mass, length, and time.

In a similar way we can show that the unit current in the electromagnetic system is an absolute unit. In that system the unit current is defined by the help of the magnetic field which it sets up, and is a quantity of the same dimensions as the strength of the magnetic shell with which it is compared in the definition. Now the magnetic potential in the field of a magnetic shell, and so of a current, is the product of the strength of the shell or of the current and a solid angle, and angle has no dimensions; so that the dimensions of current strength

are those of magnetic potential. The magnetic potential (§ 391) is defined by the strength of a magnetic pole divided by a length, and the definition of the strength of a pole (§ 384) involves only force and distance. Thus in this case also we are able, by successive definitions, to base the measurement of current upon the fundamental units of mass, length, and time.

More formally, we show, by an argument precisely similar to that used in the case of the electric charge, that the dimensional formula of a magnetic pole is $M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$. From this we obtain $M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}$ as the dimensional formula of magnetic potential and therefore of electric current. In this system also the measurement of electric current involves no arbitrarily chosen units except those of mass, length, and time.

500. The Ratio between the Units. — It will be observed that the dimensional formulas of current strength are not the same in the two systems. They differ by a factor LT^{-1}, the dimensional formula of velocity. This indicates that if a certain current is measured both in electrostatic and electromagnetic units, the ratio of the numbers expressing these measurements will not be a pure number, but a number depending on the units of length and time as a velocity does. We may state this relation otherwise by saying that the current strength which is the electrostatic unit is proportioned to that which is the electromagnetic unit by a quantity which is not a pure number but a number expressing a velocity.

All the other concepts which are defined and measured in the development of electric science are derived by definitions, which are the same in both systems, from the fundamental definition of electric current strength. The velocity which is the ratio between the electrostatic and electromagnetic units of current strength will therefore appear in the ratios between the units of these other quantities when measured in the two systems. It is an important constant when regarded simply as this ratio, and turns out to have great physical significance when it is shown to be a velocity equal to the velocity of light or of electric radiation in general.

By a more critical examination of the derivation of the units of current strength in the two systems it will appear that the occurrence of a velocity in the ratio between them is due to the fact that the explicit introduction of certain factors has been suppressed or avoided by the peculiar definitions of those units. Thus, in defining the electrostatic unit of current, we began with the electrostatic unit of charge, defined by the force between two charges in vacuum. Now the general expression for the force between two charges in any medium contains another factor, which we have not explicitly introduced in the formula defining the unit, namely, the dielectric constant K of the medium. The full formula for the force is $F = \frac{ee'}{Kr^2}$. In defining the unit charge a particular medium, vacuum, is chosen, and its dielectric constant is arbitrarily set equal to unity, so that the

K disappears from the formula. The dimensions of the dielectric constant, however, do not disappear from the dimensional formula, except on the arbitrary convention by which the electrostatic unit is defined. If we retain the dimensions of the dielectric constant, denoting them by K, the full dimensional formula of electric charge becomes $K^{\frac{1}{2}}M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$, and of electric current determined by the electrostatic definition, $K^{\frac{1}{2}}M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}$.

A similar discussion may be given of the electromagnetic unit of current. The full dimensional formula of the strength of a magnetic pole involves the dimensions of the magnetic permeability of the medium. Denoting this by μ, the full dimensional formula of the strength of a pole becomes $\mu^{\frac{1}{2}}M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$. The magnetic potential at a point whose distance from a pole m is r, in a medium whose permeability is μ, is $m/\mu r$, so that the full dimensional formula of magnetic potential, and so of current strength, determined by the electromagnetic definition, is $\mu^{-\frac{1}{2}}M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}$.

In these complete dimensional formulas the dimensions of current must be the same for both, so that the ratio between the dimensional formulas is a quantity having no dimensions, or unity. We thus get $K^{\frac{1}{2}}\mu^{\frac{1}{2}}LT^{-1} = 1$, or $K^{\frac{1}{2}}\mu^{\frac{1}{2}} = L^{-1}T$. We cannot obtain the dimensions in mass, length, and time of either the dielectric constant or the magnetic permeability, but the reciprocal of the square root of their product has the dimensional formula of a velocity. It is this quantity which is the ratio that appears in the comparison of units in the two systems.

501. Determination of v. — The velocity which appears in the ratios between the electrostatic and the electromagnetic units is commonly designated by v.

It was first determined in 1856 by Weber and Kohlrausch. These observers measured the charge of a Leyden jar in electrostatic units, and then, by discharging the jar through a galvanometer, they measured the same charge in electromagnetic units. The ratio found by them for the numbers expressing the same charge in the two units was 3.11×10^{10} centimeters per second. The ratio v expresses the number of electrostatic units of charge in one electromagnetic unit.

Lord Kelvin in 1869 determined this ratio by measuring a difference of potential in the two systems. A constant battery was joined in circuit with a conductor of very high known resistance, and the current in the circuit was measured in electromagnetic units. From Ohm's law the difference of potential in electromagnetic units was therefore given by the product of this current strength and the resistance. This same difference of potential was also measured in electrostatic units by the absolute electrometer. The ratio of the numbers thus obtained gave the value 2.82×10^{10} centimeters per second for v. The dimensions of difference of potential are determined in both systems from the relation that the product of the difference of potential and of electric charge are

equal to work done; so that the ratio of the dimensions of difference of potential in the two systems is the inverse ratio of the dimensions of charge. There are, therefore, v electromagnetic units of difference of potential or of electromotive force in one electrostatic unit. Measurements of the same ratio have been made many times by these and other methods. As greater precision has been attained in the measurements, a very satisfactory accord has been reached in the results. The value for v which is now generally adopted as being sufficiently correct for all present needs in theory and in calculation is 3.00×10^{10} centimeters per second.

EXAMPLES, XXVI

1. *To find the distribution of current in two conductors of resistance r_1 and r_2 which are joined together at their ends, and form part of a circuit.*

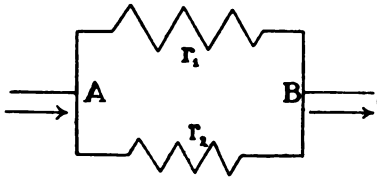


Fig. 274.

Call the potentials at the ends A and B (Fig. 274), V_1 and V_2 respectively. Then by Ohm's law the current in the upper branch is $i_1 = (V_1 - V_2)/r_1$, and that in the lower branch, $i_2 = (V_1 - V_2)/r_2$.

2. *To find the resistance of a single conductor which is equivalent to the resistance of the two conductors of Example* 1.

The current in the main circuit is i, and is divided in the two branches.

Setting $i = i_1 + i_2$, and substituting the values for i_1 and i_2, we have $i = (V_1 - V_2)\,(1/r_1 + 1/r_2)$; so that if a single conductor whose resistance r is given by the formula $1/r = 1/r_1 + 1/r_2$ is substituted for the two branches, the current and difference of potential will remain unaltered.

This result can be generalized for any number of conductors *in parallel*, or which are joined together at their ends with the terminals of the main circuit.

3. *Show that when the resistance of one of the two conductors of Example* 1 *is very large, the current in it varies proportionally to the other resistance, if the total current is kept constant.*

The conditions here stated can be approximately met by having one end of the large resistance free, and sliding it along to different points on the other resistance.

Let r_2 be a very large resistance. Then $1/r = 1/r_1$ approximately, and $ir_1 = V_1 - V_2$ approximately, so that as r_1 is changed, while i remains constant, $V_1 - V_2$ changes proportionally to r_1. But $i_2 = (V_1 - V_2)/r_2$, so that i_2 changes proportionally to r_1.

4. *To divide a current into two parts in any proportion.*

The current in a divided portion of a circuit similar to that of Example 1 is divided into two parts which are inversely proportional to the resistances of the two branches of the circuit; and by properly adjusting those resistances the required division of the current can be effected.

Such a combination of two conductors is called a *shunt*, or a *shunt circuit*.

5. *To find the currents in a divided circuit as fractional parts of the total current.*

In the divided circuit (Example 2) we have $V_1 - V_2 = ir = i_1 r_1 = i_2 r_2$, and also $$r = \frac{r_1 r_2}{r_1 + r_2};$$ hence $$i_1 = i\frac{r_2}{r_1 + r_2}, \qquad i_2 = i\frac{r_1}{r_1 + r_2}.$$

6. *If voltaic cells of the same sort are introduced in series into a circuit, show how the current varies with the number of cells introduced.*

Let e represent the electromotive force of each cell, and n the number of cells, so that ne represents the electromotive force of the cells introduced. Let r represent the resistance of one cell, and s the external resistance. Then by Ohm's law $$i = \frac{ne}{nr + s} = \frac{e}{r + s/n}.$$

If r is small in comparison with s/n it may be neglected, and we have $i = ne/s$. The current increases nearly in proportion to the number of cells introduced. This is the arrangement to use to get an appreciable current in a circuit of high resistance.

If s/n is small in comparison with r, we have $i = e/r$, and the current is not appreciably increased by introducing additional cells in series.

7. *If voltaic cells of the same sort are introduced in parallel into a circuit, show how the current varies with the number of cells introduced.*

The resistance of each cell is r, and the external resistance is s; the electromotive force of one cell is e, and there are n cells. Then the resistance of the n cells in parallel is r/n (Example 2) and the current in the circuit is

$$i = \frac{e}{s + r/n} = \frac{ne}{ns + r}.$$

If ns is small in comparison with r, the current is approximately proportional to the number of cells. This is the arrangement to use to get a large current in a circuit of low resistance.

8. *Show that, if various voltaic batteries of low resistance are introduced into a circuit of high resistance, the current in each case is proportional to the electromotive force of the battery.*

Let e represent the electromotive force, r the low resistance of the battery, and s the high resistance of the external circuit. Then by Ohm's law $i = \frac{e}{r + s}$, and if r is small, $i = e/s$ approximately, and will be proportional to the electromotive force.

9. *To compare a resistance with another one which is already known.*

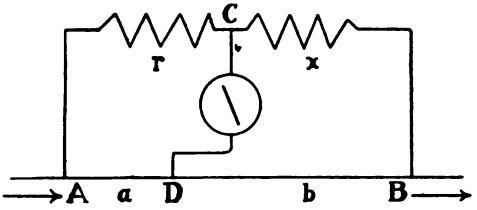


Fig. 275.

Build up a divided circuit as follows (Fig. 275): at the terminal point A of the main circuit join one end of the known resistance r, to the other end of it join an end of the unknown resistance x, and join the other end of the unknown resistance to the other terminal point B of the main circuit. Also join the points A and B by a cylindrical wire. Join one terminal of a galvanometer to the point C at which the resistances r and x are joined, and move the free terminal of the galvanometer along the wire

until a point D is found such that there is no current indicated by the galvanometer. The point D divides the wire into two lengths a and b; and the resistance x is given by the formula $x = rb/a$. This arrangement is called Wheatstone's bridge.

To prove the formula, we consider that the resistances of the two lengths of the wire are proportional to those lengths, and so may be represented by ca and cb, if c is the factor of proportion, or the resistance of unit length. Because no current passes through the galvanometer, we know that the potential at C is the same as that at D. Representing this potential by V and the potentials at A and B by V_1 and V_2 respectively and the currents in the resistances and in the wire by i_1 and i_2 respectively, we have from Ohm's law $i_1 = (V_1 - V)/r = (V - V_2)/x$; $i_2 = (V_1 - V)/ca = (V - V_2)/cb$. Eliminating the potentials from these equations by dividing one by the other, we get $x/cb = r/ca$; or $x = rb/a$.

It is evident that two sets of adjustable resistances may be substituted for the two lengths of wire, and that with the point D taken as their point of junction, changes in the resistances may be made until no current passes through the galvanometer. Representing these resistances, which take the place of ca and cb, by r_1 and r_2 respectively, we have $x = rr_2/r_1$.

502. Joule's Law. — The energy which is introduced into the circuit from the source of the current is expended in the circuit. In a homogeneous part of the circuit, as, for example, in a wire joining the poles of a battery, this energy is converted into heat. In 1842 Joule carried out a series of experiments to discover the relation between the heat developed in such a conductor and the current in it. He found that for a given conductor the heat developed is proportional to the square of the current, and that, for a constant current, the heat developed in different conductors is proportional to the resistance. This relation is known as *Joule's law*. It may be expressed by the formula

$$h = i^2 r, \tag{147}$$

in which h represents the rate at which heat is developed, the heat being measured in energy units.

Joule's law was discovered originally by experiment, but it follows immediately from the relations expressed in the formulas of § 498. For, in a homogeneous part of the circuit, in which the expended energy appears entirely as heat, the formula $h = i(V_1 - V_2)$ applies; also, since Ohm's law holds for parts of circuits as well as for the whole circuit, the formula $ir = V_1 - V_2$ also applies; and by eliminating $V_1 - V_2$ from these equations we obtain $h = i^2 r$, or Joule's law in its usual form.

503. Helmholtz's Theorem. — When we consider the whole circuit, it generally happens that some of the energy expended in it is used in doing something else than heating the conductor. It may, for example, do chemical work in electrolysis, or heat the

junction of two metals, or lift a magnet into a coil. In all such cases as these it has been found by experiment (§§ 461, 481, 492) that the rate at which energy is expended in doing this extra or local work is exactly proportional to the current. We may therefore express the relation between the total energy expended by the source of the current and the work that is done throughout the circuit, both in the form of heat and in extra work, by the equation $ie = i^2r + ia$, in which the factor a expresses the rate at which this extra work is done by a unit of current. By transposing, we obtain the current strength in the circuit expressed by the formula

$$i = \frac{e - a}{r}. \tag{148}$$

The numerator $e - a$ is evidently an electromotive force, and the formula shows that when work is done in the circuit, in any other way than in heating the circuit, the electromotive force in the circuit is less than that introduced by the source of the current, by an amount which depends on the rate at which the extra work is being done. If the electromotive force e is suppressed, leaving the conditions in the circuit otherwise the same, the electromotive force a will exist in the circuit in the opposite sense to that of e, and a current will also exist in the circuit in the sense of the electromotive force a so long as the conditions are maintained to which that electromotive force is due.

We call the electromotive force a, which is developed by doing extra or local work in a part of the circuit, the *counter-electromotive force*. By this theorem we can explain the various modes of producing the electric current. Thus when the current moves a magnet, a counter-electromotive force is set up in the circuit. A similar movement of the magnet, due to any outside action, will cause the same electromotive force to arise and so will produce an induced current. This relation was first proved by Helmholtz, who used it in illustration of the principle of the conservation of energy, for it follows, as we have seen, as a consequence of that principle. Similarly, when the current decomposes an electrolyte, and liberates dissimilar ions upon the electrodes, a counter-electromotive force is set up. If these liberated ions are otherwise introduced into the electrolyte as the electrodes, the same electromotive force will arise. Such a combination is a voltaic cell. So also, if a current does work by heating the junction of two metals, a counter-electromotive force is set up. If the same junction is heated from an outside source,

the same electromotive force is set up, and we have the thermo-electric current.

504. Practical Units. — The direct measurement of current, electromotive force, and resistance, in their absolute units, is extremely difficult, and in ordinary circumstances impracticable. It has been found necessary to establish intermediate standards, which are more easily used in ordinary measurements, and to determine the values of those standards, once for all, in absolute units.

If we know the value of the horizontal intensity of the earth's magnetism, we can compare with it the magnetic force set up by a current in the tangent galvanometer, and can calculate the value of that current in electromagnetic units, when we know the dimensions of the galvanometer (§ 469). The absolute value of a current is often determined in this way in ordinary laboratory practice, but unless the galvanometer is very exactly made, and the operation with it conducted with extreme care, the value obtained cannot be depended on as accurate. Accordingly, several observers have compared the strength of a current, measured with a very exact tangent galvanometer, with the amount of silver which the same current deposits in a second. A study of these and similar observations has led the representatives of the principal nations to accept, as the standard or electromagnetic unit of current, the current which will deposit 11.18 milligrams of silver in one second. By the aid of this number a current which is measured directly by the silver which it deposits can be expressed in absolute units.

The electromagnetic unit of current is not the one which is used in ordinary practice. The practical unit of current, called the *ampere*, is equal to 10^{-1} absolute units of current.

To determine the absolute value of a resistance we must determine in absolute units the current and the electromotive force in the circuit. The method employed by the Committee of the British Association on Electrical Standards will serve as an illustration of how this may be done. The wire whose resistance was to be determined was wound on the circumference of a large circle into a coil whose ends were joined to each other. This coil was mounted so that it could rotate at a determinate rate around its vertical diameter, and a small magnetic needle was hung at the center of the coil. When the coil was turned, it cut through the lines of force of the earth's magnetic field, and so set up induced currents in itself. The electromotive force in the coil was calculated from the rate of change of the number of tubes of force encircled by the coil. The

induced currents in the coil all tended to turn the magnet in the same sense, and the strength of the magnetic field produced by them, and so the strength of the current in the coil, was determined from the deflection of the magnet. When the electromotive force and the current were both known in absolute units, the ratio between them gave the resistance of the coil in absolute units. By comparison with the resistance of this coil, a wire of known resistance was constructed as a standard, and by comparison with it standard sets of resistances have been made with which other resistances can be compared.

The electromagnetic unit of resistance thus determined is too small to be of practical use. Instead of it a practical unit is used, called the *ohm*, which was originally designed to be equal to 10^9 electromagnetic units of resistance. It has been found more convenient to define the ohm as the resistance of a column of mercury, one millimeter in cross section and 106.3 centimeters in length, and having a mass of 14.4521 grams, at the temperature of melting ice. The resistance of such a column is very nearly equal to the ohm as previously defined.

The electromotive force of a circuit may be measured by measuring the current and the resistance of the circuit in absolute units. Certain voltaic cells have been constructed whose electromotive force is very constant, and reproduced with great precision when the cells are made up according to a prescribed formula. The electromotive forces of these cells have been very carefully determined, so that they serve as intermediate standards.

The electromagnetic unit of electromotive force is too small to be of use in practice. We use instead of it a practical unit, called the *volt*, which is equal to 10^8 electromagnetic units of electromotive force. The electromotive force of the Daniell's cell is a little greater than one volt. The electromotive force of the Weston normal cell at 20° C. has been accepted by the representatives of the principal nations as equal to 1.0183 volts.

The energy expended in a circuit in one second is measured by the product of the current and the electromotive force in the circuit. When the current is the ampere and the electromotive force is the volt, the energy expended in one second, or the rate at which energy is expended, is taken as a unit rate of expenditure of energy. This unit is called the *watt*. It is equal to one joule (10^7 ergs) per second. The horse power is equal to 746 watts.

EXAMPLES, XXVII

1. *To show that the heat developed in a divided circuit of two branches is less than it would be if the currents were not distributed inversely as the resistances.*

By Joule's law we have the heat developed equal to $i_1^2 r_1 + i_2^2 r_2$. Suppose that the currents are not distributed as found in Examples, XXVI, 1. We then may set

$$i_1 = \frac{V_1 - V_2}{r_1} + x, \qquad i_2 = \frac{V_1 - V_2}{r_2} - x,$$

if x is an unknown quantity of current. This specification conforms to the condition $i = i_1 + i_2$. Introducing these values in the expression for the heat developed, we get after reduction,

$$(V_1 - V_2)^2 \left(\frac{1}{r_1} + \frac{1}{r_2}\right) + x^2 (r_1 + r_2).$$

This is a minimum when $x = 0$, or when the currents are distributed in accordance with Ohm's law. If we set $1/r_1 + 1/r_2 = 1/r$, we have the heat developed equal to $(V_1 - V_2)^2/r$. This is the heat that would be developed in a circuit of resistance r by a current $i = (V_1 - V_2)/r$, which shows that r is the resistance of a conductor equivalent to the divided circuit.

2. *To calculate the electromotive force of the induced current produced by the motion of a circuit in a magnetic field.*

Let us suppose a circuit set up in a magnetic field so that it encloses N tubes of magnetic force of the field. If a current i is in the conductor, its energy in the field will be iN, and if the current is flowing so that its own tubes of force are in general in the same direction as those of the field, its energy will be negative and the circuit will move so as to increase the number of tubes of force enclosed by it (§ 474). Let this increase in the short time Δt be represented by ΔN. Then the increase in the negative energy will be $i\Delta N$, and its rate of increase $i\Delta N/\Delta t$. But this rate of increase of the negative potential energy of the circuit measures the rate at which work is done on it during its motion in the field, and hence the coefficient $\Delta N/\Delta t$ is the a of Helmholtz's theorem (§ 503). This being so, $-\Delta N/\Delta t$ represents the electromotive force induced, the minus sign indicating that the current due to this electromotive force is opposite to the current to which the motion is due. The induced current will therefore be one which will set up tubes of force which pass through the circuit in the opposite direction to those of the original current in it, and in the case supposed will be such as to tend to oppose the increase in the number of tubes of force enclosed by the circuit which occurs in consequence of the motion. An independent motion of a circuit in such a sense as to increase the number of tubes of force enclosed by it will induce a current which will set up tubes of force through it in general opposite in direction to those of the field. If the current in the circuit is such that the tubes of force of the field and those of the current in general pass through the circuit in opposite directions, so that the potential energy of the circuit is positive, the spontaneous motion of the circuit is such as to diminish the number of tubes of force of the field which pass through it, and the rate of change of the number of tubes of force which pass through the circuit is $-\Delta N/\Delta t$. This is the quantity a of Helmholtz's formula, and the induced electromotive force is therefore $\Delta N/\Delta t$. The plus sign indicates that the induced electromotive force is in the same sense as that of the current to which the spontaneous motion of the circuit is due. The

induced current will therefore be one which will set up tubes of force which will pass through the circuit in the same sense as those of the original current in it; and in the case supposed will be such as to tend to oppose the decrease in the number of tubes of force enclosed by the circuit which occurs in consequence of the motion. An independent motion of a circuit in such a sense as to decrease the number of tubes of force enclosed by it will induce a current which will set up tubes of force through it in general similar in direction to those of the field.

3. *To show that the time integral of the current induced by the motion of a conductor in a magnetic field is independent of the rate at which the motion takes place.*

By the time integral of the current is meant the sum represented by the formula Σir, in which r represents a very short interval of time, and i the current in the circuit during that interval. The summation is taken over the time during which the current flows. When the current strength is constant this sum is the product of the current strength and the time during which the current flows. It is equal to the quantity of electricity which flows around the circuit.

We shall assume that the motion is such that the rate of change of the number of tubes of force which pass through the circuit is uniform. Then if the number of tubes which pass through the circuit at the beginning and end of the time t is represented by N_1 and N_2 respectively, the rate at which the number changes, and so the electromotive force in the circuit, is $(N_2 - N_1)/t$. If the resistance of the circuit is r, the current strength i is given by $i = (N_2 - N_1)/rt$. Since i is constant if the electromotive force is, the time integral of the current is $it = (N_2 - N_1)/r$, and is independent of the rate at which the motion takes place.

4. *To find the time integral of the current induced by the rotation of a plane circuit in a uniform magnetic field.*

Suppose the plane circuit set up so that the tubes of force of the field are perpendicular to its plane, and then to be rotated around an axis in its own plane. At the start it encloses N tubes of force. When its plane is parallel with the tubes of force of the field it encloses no tubes, so that the time integral of the current induced by a quarter revolution is N/r. The same current will be induced by each successive quarter revolution.

5. *To find how the time integral of the current induced by rotation of a circular circuit in a uniform magnetic field depends on the radius of the circuit.*

Let R represent the number of tubes of force that pass through unit area, a the radius of the circular circuit, and ρ the resistance of unit length of the circuit. Then the area of the circuit is πa^2, so that the number of tubes of force which it encloses when its plane is perpendicular to the tubes of force is $N = \pi a^2 R$. The resistance of the circle is $2\pi a\rho$. The time integral of the current induced by a quarter revolution is therefore (Example 4)

$$\frac{N}{r} = \frac{\pi a^2 R}{2\pi a \rho} = \frac{aR}{2\rho},$$

or is directly as the radius.

6. *A circuit is made by coiling a wire several times in a narrow groove in the edge of a circular ring. Show that the time integral of the current induced in this circuit when it is rotated in a uniform magnetic field is independent of the number of turns of wire.*

Suppose there are n turns of wire. The total change in the number of tubes of force which are enclosed by the circuit, when it makes a quarter revolution, will be nN. If r represents the resistance of one turn of wire, the total resistance

is nr. Hence the time integral of the current will be nN/nr for each quarter revolution and will be independent of the number of turns of wire.

7. *Describe the magnetic field set up by the current induced in a circular circuit by a uniform rotation in a uniform magnetic field.*

Let us suppose that we are looking from above along the axis of rotation, and that the circuit rotates counterclockwise as seen from above (Fig. 276, a). If α represents the angle made by the normal to the circuit in any position with the normal when the plane of the circuit is perpendicular to the tubes of force of the field, and if A represents the area of the circle, the projection of the circle on the plane perpendicular to the tubes of force will be $A \cos \alpha$. If R represents the number of tubes of force which pass through unit area, the number of tubes which will be enclosed by the circuit in any position will be $RA \cos \alpha$, and if α changes uniformly, this number will change as $\cos \alpha$ changes when α changes uniformly. Without working out the precise law of this change, it is obvious from a table of cosines, or from a diagram, that the cosine of an angle changes infinitely slowly when the angle is zero, and only slowly when the angle is small; that the rate of change increases as the angle increases, and is greatest when the angle is a right angle. It appears, therefore, that the electromotive force induced by the rotation will be zero as the circuit passes through the plane perpendicular to the tubes of force, and will gradually rise to a maximum as the circuit completes a quarter revolution. The induced current will follow the same law of variation, and its magnetic field will be greatest when its tubes of force, which are always at right angles to the plane of the circuit at points in that plane, are perpendicular to the tubes of force of the field.

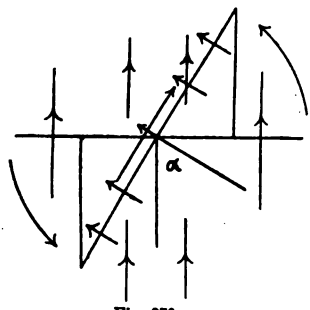


Fig. 276, a

If the tubes of force of the field are directed as indicated in the diagram, and the circuit is turning as indicated, the number of tubes of force which are enclosed by the circuit is diminishing. The induced current is such as to oppose this diminution in the number of tubes, and therefore, in the upper part of the circuit, it is directed as indicated, so that the tubes of force in the plane of the circuit are directed in general toward the left.

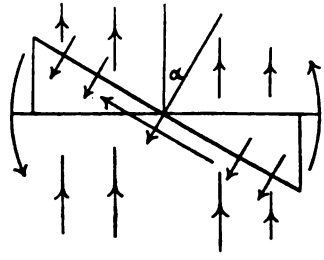


Fig. 276, b.

After a quarter revolution is completed further rotation increases the number of tubes of force which pass through the circuit (Fig. 276, b). In this part of the rotation the induced current is such as to oppose the increase in the number of tubes, and therefore in the upper part of the circuit it is directed as indicated, so that the tubes of force in the plane of the circuit are still directed in general toward the left.

The effect of each subsequent rotation will be a repetition of this. In general, then, the magnetic field of such an induced current will always be directed to the same side of a vertical plane containing the tubes of force of the field, and a

magnetic needle suspended at the center of the circuit will be acted on by a couple of varying intensity, which will always deflect it in the same sense from its original direction. When the circuit is rotated rapidly and the needle is heavy, the variations of the couple will not be indicated by the needle. It will assume a constant deflection that will depend upon the average value of the couple. (Compare with the British Association method of determining the unit of resistance, § 504.)

8. *A pair of very long parallel heavy rails joined at one end by a crosspiece is placed in a magnetic field of strength H so that its plane is perpendicular to the tubes of force of the field. The resistance of these rails and the crosspiece is supposed to be negligibly small. A straight conductor of resistance r is laid transversely over the rails so as to form a closed circuit (Fig. 277). Find the velocity with which the transverse conductor must be moved along the rails to produce an induced current of unit strength in the circuit.*

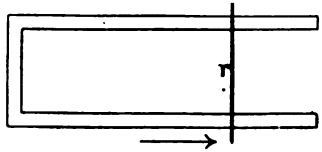


Fig. 277.

From the conditions, H tubes of force pass through each unit of area enclosed by the circuit. If the distance between the rails is d, and the transverse conductor is moved with uniform velocity v along the rails, the area swept over by it in unit time will be dv, and the increase in the number of tubes of force enclosed by the circuit in unit time will be Hdv. But this quantity also measures the induced electromotive force, so that the induced current $i = Hdv/r$; and that i may be equal to 1, we must have $v = r/Hd$.

If the distance between the rails is unit distance and if the field is of unit strength, the numerical value of the velocity measures the resistance.

9. *To find the time integral of the current induced in a plane circuit by a magnetic pole which traverses a closed path linked with the circuit.*

Suppose the pole of strength m placed at the point a in the plane of the circuit and outside of it (Fig. 278), so that no tubes of force of the pole pass through the circuit. When the pole is carried on the path indicated to the point b, one-half of its $4\pi m$ tubes of force have passed through the circuit, and a total current equal to $2\pi m/r$ has been induced in the circuit in such a sense as to oppose the increase in the tubes of force passing through it. Looking down on the circuit, the current is in the direction indicated. When the pole moves on back to the point a, one-half of its $4\pi m$ tubes of force are removed from the circuit, and a total current equal to $2\pi m/r$ is again induced in the circuit. The sense of the current is such as to oppose the removal of the tubes of force, and is the same as in the first part of the motion. The total current is therefore $4\pi m/r$, and is all in the same sense.

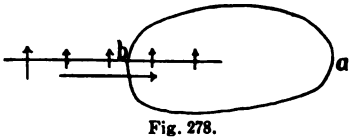


Fig. 278.

10. *To find time integrals of the currents induced when a magnet is thrust through a circuit.*

There are $4\pi m$ tubes of force proceeding from the north pole of the magnet to the south pole (§ 397). When the magnet is thrust through the circuit until its equatorial plane is coincident with the plane of the circuit, these $4\pi m$ tubes pass through the circuit and the time integral of the current which is induced

in it is $4\pi m/r$, directed in such a sense that the tubes of force of the current are oppositely directed to those of the magnet. When the magnet is thrust farther on and away from the circuit, $4\pi m$ tubes of force are removed from the circuit, and the time integral of the current which is induced is again $4\pi m/r$. The direction of the tubes of force of the magnet is the same as in the first half of the motion and the direction of the current is therefore reversed.

11. *To find the average electromotive force induced by closing and opening a circuit containing an electromotive force.*

If we represent the final current in the closed circuit by i, the number of tubes of force enclosed by it will be Li (§ 471). On closing the circuit the number of tubes enclosed by it rises in the very short time t from zero to this final value. The average electromotive force induced is therefore Li/t, and it is in the opposite sense to the electromotive force of the circuit. The current therefore rises gradually and not instantaneously to its full value.

On opening the circuit, the number of tubes of force enclosed by it falls in the very short time t' from its initial value to zero. The average electromotive force induced is therefore Li/t', and it is in the same sense as the electromotive force of the current. The current after opening the circuit is therefore temporarily continued, and under some conditions is increased. Because of the shortness of the time t' the induced electromotive force may be high. It gives rise to the extra current (§ 476).

505. Theories of the Electric Current. — From the first the electric current was thought of as a continuous transfer of electricity around the circuit. In the one-fluid theory this transfer was all in one direction; in the two-fluid theory positive electricity was supposed to move in one direction, and an equal amount of negative electricity in the opposite direction. This simple conception was consistent with the laws of electrolytic action, as well as with the general course of thought which led to the discovery of the electric current. It was not, however, sufficient to account for the actions of currents on each other, which plainly do not follow the simple laws of attraction and repulsion of static charges.

The problem of assigning such properties to the moving charges which were assumed to constitute the current as would account for the observed laws of the interaction of currents was at last taken up by Weber, who undertook to explain the actions of currents by the hypothesis that moving charges act on one another, not only with their electrostatic forces, but also with other forces which depend upon their velocities. By developing this hypothesis, Weber was able to account for the actions of steady currents on each other. By making the further hypothesis that additional forces arise when the velocities of the moving charges are changing, Weber was also able to account for the induction of currents.

The theory of Weber assumed that the actions between currents

take place in some direct manner, which does not depend in any way on the medium between the currents, and which depends only on the distance between them. It may therefore be called a theory of electric action at a distance. The researches in electricity and magnetism in which Faraday was for many years engaged convinced him that electric and magnetic actions are not simply actions at a distance, but take place by means of the intermediate actions of some medium. By reflecting on these views of Faraday, Maxwell was led to attempt to give them mathemetical form, and so to construct a theory of what we may call action through a medium. One form of his theory, in which he describes a medium which will account for electrostatic action, has already been described. An essential feature of this particular form of the theory, and indeed of Maxwell's theory in all its forms, is that there can be no such thing as an open circuit, that is, there can be no flow of electricity which begins and ends in a limited conductor. According to Maxwell's theory, the flow of electricity in a limited conductor is accompanied by an electric displacement in the dielectric around the conductor, so that the circuit is completed through the dielectric. While this displacement is being set up in the dielectric, it is so far like a current that it produces the same magnetic field that a current of the same strength and similarly distributed would produce.

In the most general form of Maxwell's theory, the hypothesis is made that the ether and the electricity, which is distributed everywhere in it and which can be displaced in a dielectric and move freely in a conductor, conform to the general principles of mechanics, in so far that if, to the quantities which enter into the general equations of dynamics, there is given an electric interpretation, those equations will represent the mode of action of the electric and magnetic fields. By the development of this hypothesis Maxwell was able to show that it leads first to the induction of currents, then to the actions of steady currents, and finally to the actions of electricity in equilibrium.

If we adopt Maxwell's theory, we conceive the act of setting up a current to involve not only starting a stream of electricity along the conductor, but also setting in action — we might by analogy say, in motion — the ethereal mechanism in the conductor and in the dielectric surrounding the conductor.

506. Poynting's Theorem. — When this mechanism is in action, the region occupied by it becomes the magnetic field of the con-

ductor. To set up this field requires energy, and since the energy from the source is supplied only at a certain rate, and is at first divided between that expended in the circuit and that spent in setting up the magnetic field, the current in the conductor rises only gradually to its full value. When it is fully established the maintenance of the action in the magnetic field requires no more energy. According to the mode of representation adopted by Poynting, the flow of energy from the source into the dielectric does not stop when this steady state is reached, but is continued at a constant rate, and the energy, passing through the dielectric in a certain definite manner, reaches the conductor by way of the dielectric and is transformed in it. When the circuit is broken and the flow of electricity ceases, the conditions which sustain the action of the mechanism cease also, and the energy stored up in the dielectric leaves it and appears in the conductor as the energy of the extra current.

507. Alternating Current. — When the current is steady, Poynting's theorem shows that the expenditure of energy in an ordinary homogeneous part of the circuit, resulting in the development of heat in it, will be uniform across the whole cross section of the conductor. But this will not be the case when the current varies periodically. A current which so varies may be developed by a properly constructed dynamo machine, which is arranged to set up an electromotive force in the circuit varying continuously and periodically between two extreme and oppositely directed values. Such a current is called an *alternating current.* It was shown by Heaviside that the energy which enters the conductor from the dielectric, when the current in it is alternating, is transformed into heat before the current distributes itself uniformly over the whole cross section of the conductor, so that the principal development of heat is in the outer layers of the conductor. This effect has been demonstrated by experiment. The extent to which the current penetrates the conductor depends on the period of the alternations, being greater as this is longer. For very rapid alternations, the current is confined almost entirely to the surface of the conductor.

This action of alternating currents is a consequence of Maxwell's theory, and may be considered a verification of it.

Still, Maxwell's theory, if it had gone no further than this, would have been little more than a simple alternative to the theory of Weber, to be adopted by those who prefer to think in terms of

action through a medium rather than in terms of action at a distance.

Maxwell, however, was able to draw from this theory another more important series of consequences, which do not follow from Weber's theory, and which were subsequently verified by experiment. We shall now turn our attention to this portion of Maxwell's theory and to its experimental demonstration.

CHAPTER XXI

ELECTRICITY AND THE ETHER

508. Electromagnetic Waves. — It has already been described how the flow of electricity in a limited conductor, on Maxwell's theory, is accompanied by an electric displacement in the dielectric. If this flow alternates in direction, or is oscillatory, the displacement in the dielectric will undergo similar oscillations. It follows from Maxwell's theory that when such oscillations are set up, they will proceed outward from their origin, as electromagnetic waves, with a definite velocity. In the ether, according to the theory, this velocity is that which expresses the ratio between the electrostatic and the electromagnetic units. In other bodies it is equal to this velocity divided by the square roots of their dielectric constants.

To appreciate the mode of progress of these electromagnetic waves, let us fix our attention on a portion of the ether through which such waves are passing. The waves may be thought of as being of the simplest type, so that the movements of the electric charges which undergo displacement are simple harmonic. As these charges move they set up magnetic fields around themselves similar to those which would be set up around elements of current. The strengths of these fields therefore undergo simple harmonic changes of the same period as that of the varying displacements of the charges, having their maximum values when the charges are passing with their maximum velocity through the positions of equilibrium which they would occupy in the undisturbed condition of the ether, and being zero when the charges are momentarily at rest in their extreme displacements. The varying magnetic fields act inductively on the next contiguous set of charges, and set them in motion, according to the laws of the production of induced currents. These moving charges in turn set up magnetic fields, and the action described is repeated in successive portions of the ether so that the original disturbance travels out as an electromagnetic wave. If the electric forces which set up the displacements are in some one direction, the magnetic forces which are set up

by the movements of the charges are perpendicular to that direction. The waves in this case are polarized, and consist essentially of two mutually perpendicular elements, — the electric force, with the displacement of the charges which corresponds to it, and the magnetic force which arises from the motions of the charges. The electric force, or, more exactly, the electric induction, follows the laws of the luminous vibrations in Fresnel's theory of polarized light; the magnetic force follows the laws of the luminous vibrations in the Neumann-MacCullagh theory (§ 359).

To describe the transmission of electromagnetic waves more specifically let us consider a plane polarized electromagnetic wave. Let the points a, b, . . . n (Fig. 279) represent points at which electric charges are at rest in the undisturbed condition of the ether, and let us suppose that these charges are all similarly displaced perpendicularly to the plane of the paper and are vibrating in simple harmonic motions in the lines of their displacements. Let us suppose that the charges are so moving, at some particular instant, as to be equivalent to currents directed downward through the plane of the paper. The magnetic forces set up around them will then be directed as indicated by the curved arrows in the diagram. The movements downward of the charges will thus set up near the charges what is equivalent to a line of magnetic force of varying strength, directed toward the left. This varying magnetic force will produce the equivalent of an induced current in the neighboring ether, and hence will set up an electric force above the line in the diagram in the opposite direction to that in which the charges are moving and so adapted as to bring them to rest; while its effect on the charges at the points a', b', . . . n' below the line will be to set them in motion in the same direction as that in which the charges at the points a, b, . . . n were originally moving. The condition which originally prevailed in the first line of charges will thus be transferred to the second line, and in a similar manner to successive lines, so that the original disturbance will be transmitted through the ether in a direction perpendicular to the planes in which the electric and magnetic disturbances were set up.

Fig. 279.

The supposed electric charges are so near together that when they move in the same sense and with the same velocity, the effect they produce is that of a plane current or current sheet of periodically varying strength. The action just described may then be

stated more concisely by saying that this varying current sheet sets up ahead of it a plane sheet of varying magnetic force, which in turn sets up a current sheet behind itself which annuls the original current sheet, and a current sheet ahead of itself which is a reproduction of the original current sheet.

509. Electromagnetic Theory of Light. — In all their essential characteristics these waves of electric and magnetic force are exactly like waves of light. In default of direct experimental evidence, Maxwell ventured to assume that the waves of light are electromagnetic waves, and to examine their properties for a confirmation of his theory. A comparison of the velocity which is the ratio between the two systems of units with the velocity of light showed that the two velocities are very nearly equal, and more accurate determinations of both these quantities than Maxwell had at his disposal have since shown that they are really equal within the limits of probable error. The velocity of light in any other body than ether is equal to its velocity in the ether divided by its index of refraction. If, therefore, light waves are electromagnetic waves, the index of refraction of a substance ought to be equal to the square root of its dielectric constant. In the case of paraffine, which was the only body for which he knew both the index of refraction and the dielectric constant, Maxwell found that this relation nearly, though not exactly, held true. Later investigation has shown that it rarely holds true with light waves; but an explanation of this circumstance can be given which permits us still to regard light waves as electromagnetic waves. In addition, according to the theory, good conductors ought to be opaque to light and poor conductors transparent, for when the electric disturbance in the wave falls upon a good conductor, it should set up currents and its energy should be transformed into heat, while in a poor conductor it should continue without any such transformation. Maxwell found, in fact, that this relation generally holds true, although the degree of transparency exhibited by the conducting metals when in thin sheets was greater than he had expected. Later observation, and an extension of the theory, have shown that the facts of observation confirm the theory.

Maxwell was able to go no further than this, and the *electromagnetic theory of light*, as it was called, remained an interesting speculation, with some evidence in its favor, until the year 1888, when the study by Hertz of electromagnetic waves and their properties confirmed the general theory of Maxwell in the most complete manner.

510. Hertz's Experiments. — To set up electromagnetic waves, Hertz used the electric spark passing between two knobs joined by short rods to two similar metal plates standing in the same plane. This arrangement is called the *vibrator*. The discharge in such an electric spark is oscillatory. It is not the passage of electricity in one single leap across the gap, but is instead a succession of passages of electricity, alternately in opposite directions and in diminishing quantity, until equilibrium is attained. These alternate passages occur at equal intervals of time, and are therefore adapted to serve as the origin of a short train of waves.

Hertz detected the passage of these waves by using a wire rectangle or circle, in which a small gap was opened so that, if a current was set up in the circuit, its presence could be detected by a spark at the gap. He constructed this circuit of such a length that the period of the free oscillation of electricity in it was the same as that of the discharge. The electric impulses coming to it in the wave were therefore synchronous with its own free vibration, and thus the electric effects in it were heightened. This instrument was called the *resonator*. After showing that he could detect an electric disturbance, which was presumably an electric wave, whenever the spark passed at the vibrator, Hertz used the resonator to compare the velocity of electric waves in a wire with their velocity in air. The results of these experiments were inconsistent with Maxwell's theory, but Hertz afterwards recognized that there was probably some undiscovered error in the way in which the experiments were made, and the observations of others who have tried similar experiments confirm the conclusions of the theory, that the velocity of the electromagnetic waves in the wire is the same as that in the air.

According to the theory an electromagnetic wave is reflected with reversal of phase of the electric force at a conducting surface on which it falls perpendicularly. Hertz allowed the waves to fall on a large sheet of zinc and observed that the resonator behaved as if nodes of electric force, due to the combination of advancing and returning waves, were formed at the surface and at points in front of the surface which were taken to be half a wave length apart. From these observations and the calculated period of the vibrator the velocity of the waves was calculated to be of the same order of magnitude as the velocity of light.

Sarasin and De la Rive subsequently showed that the positions of these apparent nodes depended upon the free period of the resonator and not upon the period of the vibrator. The phenomenon is

now explained as due to the combination in the resonator, in which the vibrations set up by a wave persist for some time, of one of the oscillations in it with an impulse returned to it by reflection. The velocity of the waves, calculated on this explanation, will be the same as that obtained from the other. The velocity of light and the velocity of the electromagnetic waves are the same.

Hertz constructed a large prism of pitch and with it found that the electromagnetic waves are refracted like light waves. The index of refraction which he determined was of the same order of magnitude as that obtained for similar substances with light.

The disturbances in the spark at which the electromagnetic waves originate are all in one line, and so the waves which come out from the origin are polarized. Hertz proved this to be the case by interposing in the path of the wave a screen made of a large number of parallel wires set at small distances from one another. When these wires were set parallel with the direction of the spark they acted as a conducting surface to absorb or reflect the vibrations. When they were set perpendicularly to the vibrations, their effect was not nearly so marked, and the waves passed through the screen. The polarization of the electromagnetic waves, and the complete analogy between their behavior and that of light waves, was shown by the experiments of Trouton. Trouton allowed electromagnetic waves to fall obliquely upon a thick stone wall, and observed the reflected and refracted waves. He found in general that the relative intensities of the reflected and refracted waves depended upon the angle of incidence, and also upon the angle between the direction of the electric vibrations and the plane of incidence, and that, in particular, when the electric vibration was in the plane of incidence, and the angle of incidence had a certain value, corresponding to the polarizing angle of light, the reflected wave entirely disappeared. In fact, the behavior of these waves was exactly like that of polarized light under the same conditions, according to Fresnel's theory.

Numerous experiments have been made with electromagnetic waves to test Maxwell's relation between the dielectric constant of a substance and its index of refraction. The theory has been fully confirmed. The reason why light waves do not generally show an agreement with the theory seems to be that the dielectric constant is not the same for the rapidly alternating electromotive forces of the light waves as it is for the slower alternations by which alone we can determine it.

511. The Ether. — By these and many similar experiments the properties of the electromagnetic waves have been shown to agree precisely with those deduced from the theory of Maxwell. We may therefore consider this theory as confirmed, at least in its essential features. We may consider it proved that electric action takes place in a universal medium, which is essentially the same everywhere, except in so far as its properties are modified in particular places by the material bodies there present. From the general agreement in behavior between electromagnetic waves and light waves, it may also be taken as proved that this medium is the ether, with the properties of which we have to some extent become acquainted in our study of light. We may therefore adopt the electromagnetic theory of light without reserve, and consider the subjects of magnetism, electricity, and light as forming parts of a general science of the ether.

512. Pressure of Light. — A remarkable conclusion of the electromagnetic theory is that light will exert a pressure against a surface upon which it falls. This conclusion was unconfirmed for many years. An indirect proof of it was given by Boltzmann, who showed that if it were correct, it could be proved, by the methods of thermodynamics, that the rate of radiation from a hot body should be proportional to the fourth power of the absolute temperature. Now this law of radiation had been announced by Stefan (§ 370) as the result of his examination of the experimental facts connected with radiation; so that this fact, at least, supported Maxwell's conclusion. More recently Lebedew, and E. F. Nichols and Hull, have shown, by direct observation, that light does exert a pressure such as the theory describes, and have measured its magnitude and found it to conform precisely to that predicted by the theory.

513. Magnetic Effect of Electric Convection. — The conception that an electric current consists of a stream of electricity along the conductor leads to the hypothesis that an electric charge, when carried rapidly along, will set up a magnetic field. This hypothesis was confirmed by the experiments of Rowland, who observed a magnetic field in the neighborhood of a charged disk kept in rapid rotation. Variants of Rowland's experiment, executed by other observers, have yielded similar results.

CHAPTER XXII

ELECTRICITY AND MATTER

The Electric Discharge

514. The Electric Spark. — A critical examination of the spark passing between oppositely charged bodies shows certain peculiarities in it, which of themselves would lead to some modification of the theory of electricity which has hitherto been presented. As the discharge by means of the spark is variously modified, these peculiarities become more and more prominent, and the necessity of some development of the theory of electricity becomes more and more apparent. The most important advance that has been made in physical science in recent years began with the study of these peculiarities of the electric discharge.

The *spark*, as ordinarily seen, is a bright line or band of light, nearly straight when it is short, but broken or zigzag if its length is greater than two or three centimeters. By observations of the spark in a rapidly revolving mirror, Feddersen showed that in ordinary circumstances it consists of a succession of sparks occurring at regular intervals, and gradually diminishing in intensity. The whole spark, however, lasts for so short a time that, for many purposes, we may consider it instantaneous. The special peculiarity of the spark which was referred to in the preceding paragraph is that the two ends of it, when carefully examined, do not seem to be exactly alike. There is, in fact, a characteristic appearance at the positive electrode which differs from that at the negative electrode. This difference between the two ends of the spark becomes more apparent when the discharge is modified, by proper manipulation, into what is known as the *brush discharge*. The brush discharge occurs between electrodes which are at a considerable distance, perhaps 20 centimeters or so, from each other. When it is established, a narrow band or trunk of light originates at the positive electrode, and branches out into innumerable fine lines of light, which cease to be visible before they reach the negative electrode. At the negative electrode there may generally be seen one or more short brushes of light extending from the electrode

for not more than a centimeter. Between these and the positive brush there is no visible evidence of the discharge at all. To ordinary observation, at least, the brush discharge seems to be continuous.

By suitable manipulation the discharge may be made to take still another form, called the *glow discharge*, in which the characteristic difference between the positive and negative ends of the discharge is also apparent. In this form of the discharge a faint luminous glow appears in a thin layer over the positive electrode. Negative brushes, similar to those already described, generally appear at the other electrode. These may sometimes be made to coalesce so as to form what is called the negative glow. Even when this is done, however, the appearance of the negative glow is distinctly different from that of the positive one. Except for these glows, no light is produced by the discharge. The discharge in this case also seems to be continuous.

In all our previous discussions we have treated the vitreous and resinous electricities, or the positive and negative electric conditions, as if they were the precise counterparts of each other, and differed in their properties only by being opposite to each other. In these various forms of discharge we have found differences between the conditions at the two electrodes which indicate that the differences between the two kinds of electricity may involve other features than those which have so far been assumed. We shall find that this conclusion is confirmed, and our knowledge of the peculiar characteristics of the two kinds of electricity very much enlarged, by the further study of the discharge.

515. Discharge in Low Vacua. — For the study of the discharge in rarefied gases we use what is called a *vacuum tube*, that is, a glass tube or bulb from which the air or other gas or vapor which has filled it can be withdrawn by means of an air pump, and which is furnished with two terminals or electrodes supported on platinum wires, generally sealed into the walls of the tube. In future we shall call the electrode which is joined to the positive pole of the machine the anode, and that joined to the negative pole of the machine the cathode.

When the machine is in operation and the air is gradually withdrawn from the tube, the discharge, which is at first an intermittent spark, becomes more and more frequent as the vacuum improves, until it is appreciably continuous. In this condition almost the whole of the interior of the tube becomes luminous with a faint

rosy light. On examining the two electrodes we perceive very characteristic differences in the discharge around them. It is not necessary to describe these differences in detail. It is sufficient to say that the anode is generally covered with a sheath or layer of rosy light apparently similar to that in the body of the tube. This light, which nearly fills the tube, is called the *positive column*. In very many cases it appears to be broken into saucer-shaped layers or *strata* of light, separated by darker spaces. The concave sides of these strata are turned toward the anode, and the light around the anode seems to be simply one of these strata of modified form. The positive column terminates in the tube, and is separated from the cathode by a *dark space*. The cathode is surrounded by a bluish glow, extending to a short distance in all directions from it, and separated from it by a non-luminous region or negative dark space.

The color of the positive column, and the size and distinctness of the strata, depend largely upon the gas or vapor which is present in the tube. The relative prominence of the different features of the discharge depends upon the extent to which the exhaustion of the tube is carried. As the exhaustion proceeds, the positive column diminishes in length and brightness and the negative glow moves farther away from the cathode, thus increasing the width of the dark space around it.

516. Discharge in High Vacua. — In very high vacua, such as can be obtained only by the use of the best air pumps, the region occupied by the negative dark space is so enlarged as to fill practically the whole of the tube. In this condition it may be seen that this space is not entirely dark, but is filled with a faint bluish light. This light is indicative of the presence of the so-called *cathode discharge*. The properties of the cathode discharge were very fully investigated by Crookes, and the tubes used for the investigation are frequently called Crookes' tubes.

Wherever the cathode discharge reaches the walls of the tube, it sets up a peculiar phosphorescent illumination, the color of which depends on the nature of the glass. In the soda glass ordinarily employed, the color is a pale green. Similar phosphorescence, with characteristic colors, is set up in very many natural minerals, or chemical preparations. By the aid of a screen covered with a suitable preparation the path of the cathode discharge may be studied.

By the direct examination of the light of the cathode discharge, as well as by the use of phosphorescent screens, Crookes showed that

the discharge proceeds from all parts of the cathode in lines normal to its surface. If, therefore, the cathode is a flat plate which faces down the tube, a principal part of the cathode discharge will be a nearly cylindrical column lying around the axis of the tube. If the cathode is a hollow spherical cup, with its concave side facing the tube, the discharge will converge to the center of the sphere, and will diverge in a cone after passing through that center. By interposing an obstacle, like a sheet of metal, in the path of the cathode discharge from a flat cathode, a sharply defined shadow is cast on the phosphorescent screen placed beyond the obstacle.

Crookes noticed that when the cathode discharge was directed against light films of glass, they were moved as if a blast of air had fallen on them. He constructed various mechanisms, such as wheels furnished with vanes or paddles, and found that when the cathode discharge was directed against them they were set in rotation as they would be by a blast of air. He concluded that the cathode discharge contains matter moving with a high velocity. In his view this matter consisted of the molecules of gas still remaining in the tube, which were first negatively electrified by contact with the cathode, and were then repelled from it.

Another experiment tried by Crookes was in harmony with this view. In it a thin piece of platinum foil, placed at the center of a spherical cathode, was made red-hot at the point where the cathode discharge met it. The heat thus developed can of course be explained by the impacts of the moving matter in the discharge.

When a magnet was brought near the tube, the cathode discharge seemed to be affected by it. To test this more fully, Crookes allowed the discharge from a flat cathode to pass through a narrow slit in a sheet of metal, and then along a phosphorescent screen, on which it developed a long narrow band of light. This band of light was ordinarily straight, but when a magnet was brought near the tube it was bent into a curve. The direction of curvature depended on the direction of the magnetic force which set it up, and was such as to indicate that the stream of particles was negatively electrified (Fig. 280).

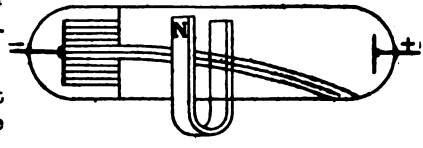


Fig. 280.

It should be said at this point that Crookes was in error in ascribing the motions of the light bodies and the heating of the platinum foil to the impacts of molecules of gas. The bodies which move from the cathode in the cathode stream are the electrons which will be described hereafter. They produce the

heat perceived in the platinum foil by their impact, and also heat the light bodies against which they fall; but their momentum is not sufficient to explain the motions of those bodies. These are probably due to an action in the remaining gas in the tube, similar to that observed in the radiometer (§ 187), called forth by the higher temperature developed on the sides of the bodies against which the cathode stream impinges.

517. Cathode Discharge Composed of Negatively Charged Particles. — The conclusion that the discharge from the cathode is composed of independent negatively charged particles is one of great importance. It was subsequently verified by an experiment of Perrin, which was repeated by J. J. Thomson. In this experiment the cathode discharge was allowed to pass through a small opening in one end of a brass box contained in the tube and joined with an electroscope. As the discharge continued, the box became more and more negatively electrified, as would be the case if it were receiving continual accessions of negative charge in its interior.

J. J. Thomson succeeded in showing that the discharge was also acted on by an electric field as if it were a stream of negatively charged particles. This cannot be done by a field set up between electrodes which are outside the tube, because the induced charges in the walls of the tube destroy the field within it; but by placing two flat electrodes within the tube, so that the discharge could pass between them, and by using a very highly exhausted tube, so that the residual gas in it was not enough to discharge the electrodes rapidly, Thomson was able to show that, when the electrodes were oppositely charged by being joined to the poles of a battery, the cathode discharge was deflected between them. The sense of the deflection indicated that the discharge was repelled by the negative electrode (Fig. 281).

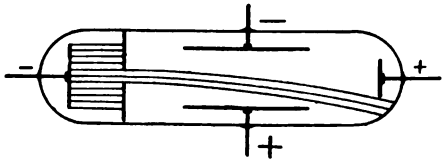


Fig. 281.

518. Constituents of the Cathode Discharge. — It was shown by Hittorf that when a magnetic field which is uniform throughout the whole length of the discharge is set up around the tube the cathode stream becomes the arc of a circle. As was shown by Schuster, this observation is consistent with the hypothesis that the cathode stream is made up of negatively charged particles moving with a high constant velocity. The amount of the deflection in a given magnetic field depends upon the velocity of the particles, and

upon the ratio of their masses to the charges which they carry. By combining the results of his own observations of the deflection in a magnetic field with those of his observations in an electric field, J. J. Thomson was able to determine the velocity of the particles in the stream, and the ratio of their masses to their charges. He found that the velocity was very great, in many cases not less than one-third the velocity of light. The ratio of the mass to the charge was about one-thousandth part of the ratio of the mass of the hydrogen atom to the ionic charge which is associated with it in electrolysis. On the supposition that the charge of each particle in the stream is equal to the ionic charge, which seems from all the phenomena of electrolysis to be a natural unit of electricity, it would follow that the mass of the particle is about the one-thousandth part of the mass of the hydrogen atom. This conclusion, which, as here stated, involves a certain element of hypothesis, was afterwards confirmed by Thomson in several ways which will subsequently be described.

519. **The Anode Discharge.** — The effects produced by the cathode discharge were so striking that for a long time they occupied every one's attention, so that the anode discharge was entirely neglected. By piercing a flat cathode, placed perpendicular to the axis of the tube, with holes, and observing behind it with a phosphorescent screen, it was found that streams or discharges also pass in the tube in the opposite direction to the cathode stream, as if they came from the anode. The effects produced on these anode discharges by the magnetic and electric fields were examined by W. Wien and by Ewers, and it was proved that the velocity of the particles in them is very much less than that of the particles in the cathode discharge, and that the ratio of the mass of these particles to their charges is of the same order of magnitude as the ratio of the mass of an atom of an ordinary metal to the ionic charge. These experiments proved at the same time that the charges of these anode streams are positive. It is surmised that these streams consist of positively charged atoms of metal torn off from the electrode.

520. **Lenard Rays.** — While experimenting with a cathode discharge received upon an aluminium screen, Hertz discovered that phosphorescent effects could be obtained behind the screen. Guided by this discovery, Lenard constructed a vacuum tube, at one end of which, opposite the flat cathode, the glass wall was pierced by a small aperture or window, closed with a sheet of aluminium foil.

When the cathode discharge was directed against this window, peculiar effects were obtained in the region immediately around it, outside the tube. Among these effects were the production of phosphorescence, and of photographic action on a photographic plate. In the experiment with the photographic plate, the plate was screened from the light of the tube by a covering of black paper, through which the action penetrated. It was found that different bodies, placed on the photographic plate, were penetrated by the action in different degrees, it being in general true that the denser bodies were less easily penetrated than those which were less dense. The cause of these actions is evidently the cathode discharge which has passed through the aluminium.

Lenard, at a later period, examined the nature of the cathode discharge with a tube of peculiar construction, which was divided into two parts by a glass wall, furnished with an aluminium window. One of these tubes was furnished with anode and cathode, and the air was exhausted from it to the particular degree at which the cathode discharge is most strongly excited. In the other tube as perfect a vacuum was made as could be made. With this arrangement the cathode stream which passed through the window proceeded in the second tube in a perfect vacuum. Lenard examined the deflections of the cathode stream, in the second tube, produced by the magnetic and electric fields, and obtained results for the velocity and for the ratio of the mass to the charge which agree fairly well with those of Thomson.

At the time Lenard discovered these effects he considered the cathode discharge to be some sort of wave disturbance or radiance in the ether, and spoke of the effects as due to rays. These supposed rays were therefore called *Lenard rays*. On the view of the nature of the cathode stream which we have adopted, and to which Lenard finally came, they are not rays in any proper sense.

521. Roentgen Rays. — A series of additional effects produced by the discharge in a vacuum tube was discovered in 1896 by Roentgen. These effects are ascribed to what, as will be seen, we may properly call *Roentgen rays*. When the cathode discharge falls on the wall of the tube, or better, when it is concentrated by a spherical cathode upon a block of platinum, the place upon which it falls becomes the source of these rays. Some of them, at least, pass out through the walls of the tube, and their effects may be studied outside of it. The most noticeable effect, and the one by which the rays were discovered, is that of affecting the photo-

graphic plate, and of exciting phosphorescence in many substances. The rays pass through various substances which are opaque to light to a degree which seems to depend, other things being equal, upon the densities of those substances. The photographic plate can therefore be exposed to the rays while it is enclosed in a plate holder. Judging from the positions of the shadows cast on the plate by various objects, Roentgen concluded that the action travels from the point of origin in straight lines; and, from the intensity of the photographic effect at different distances, he concluded that the intensity of the action is inversely as the square of the distance from the point of origin. In these respects the action is comparable with light, and we therefore designate as rays the lines along which it proceeds. In no other respects have the Roentgen rays been proved to be the same as light rays. There is no evidence of true reflection, though there is secondary radiation. Neither is there evidence of refraction or of polarization. In a recent experiment Haga and Wind have brought forward evidence which indicates a diffraction, from which, if it is to be accepted, we may conclude that the wave lengths in the rays are not more than one five-thousandth as long as ordinary light waves. The hypothesis that they are such waves is compatible with what is known about them. Even if we do not accept the evidence offered to prove that they are diffracted, the wave theory of their nature is still the most acceptable one. The waves are probably set up by the sudden stoppage of the negative charges in the cathode stream, and are therefore rather a series of irregular pulses than a regular train of waves.

522. The Negative Electron. — It was shown by J. J. Thomson that when the Roentgen rays pass through a mass of air they render it a conductor.

According to Thomson the conducting power is given to the gas by an ionization of its atoms, that is, by a separation of some of the atoms into positive and negative portions.

He undertook an investigation of the properties of ionized air in order to determine the magnitude of the charges carried by the ions. Rutherford had determined, in 1897, the velocity of the ions in an electric field of definite strength. By ionizing air between two electrodes, using the Roentgen rays for that purpose, and by measuring the current transmitted by them between the electrodes when their difference of potential was known, a quantity was obtained equal to the number of the ions, multiplied by the

charge on each, multiplied by their velocity. The velocity being known from Rutherford's experiment, the charge could be determined if the number of the ions could be determined. To accomplish this Thomson used a discovery of C. T. R. Wilson, that when dust-free air, saturated with water vapor, is ionized, and then suddenly expanded to a certain degree, less than that required to produce condensation in ordinary air, condensation of water occurs on the negative ions. By observing the rate at which the fog thus formed settled, the size of the drops was calculated, and from an observation of the quantity of water condensed, the number of drops, and so of the negative ions, was determined. From the data thus obtained it was shown that the charge of the negative ion is the same as that belonging to the hydrogen atom, and called the ionic charge.

It was shown by Elster and Geitel that when ultra-violet light falls on a charged conductor in air, the conductor is speedily discharged if its charge is negative, while it is not discharged if its charge is positive. The effect seems to be to excite conditions in which negative charge is liberated from the body, for if the light falls on an uncharged conductor, it will gradually become positively charged. This action is best obtained with clean zinc.

It was shown by the same observers that the discharge from a plate of zinc is diminished, and finally checked altogether, if the zinc is in a sufficiently strong magnetic field, the lines of force of which are parallel with its surface. The experiment succeeds only when the zinc is in a vacuum. On the supposition that the escape of the negative charge takes place by the emission of negative ions from the surface, Thomson showed that the effect of the magnetic field ought to be to change the straight paths of the ions into cycloids, and that if the discharge is taking place between the zinc and an electrode placed parallel with it and near it, a current ought to pass between the plates, of which the intensity will remain appreciably the same so long as the cycloidal paths of the ions meet the second plate, but will change abruptly, or at least very rapidly, for such a strength of the magnetic field that most of the cycloidal paths do not meet the second plate. By trial Thomson convinced himself that such an abrupt change in the current occurred, and that his hypothesis was therefore justified. From the distance between the plates he determined the magnitude of the cycloidal paths, and from this and the observed values of the difference of potential between the plates, and the strength of the magnetic field, he computed the ratio

of the mass of the ion to the charge upon it. He found this ratio to be the same as that which he had determined for the particles in the cathode stream. By condensing water drops on the ions, in the way already explained, he determined their number, and hence determined the charge on the ion, which he found in this case also to be equal to the ionic charge of the hydrogen atom. In this way, without hypothesis, he showed that the mass of the negative ion which is associated with an ionic charge is about the one-thousandth part of the mass of a hydrogen atom. It can hardly be doubted that the negative charges in the cathode stream are also ionic charges, and that the mass of the particle is far less than that of the hydrogen atom.

The negative ions thus recognized must be distinguished from the ions in electrolysis, which are masses of matter consisting of one or more atoms, and associated with a number of ionic charges equal to the valency. To denote this difference, we may call the negative ions of the cathode stream, or those liberated in a gas by the Roentgen rays, or otherwise, by the name *electrons*.

To describe these results, Thomson supposes that the atoms of a gas consist of a positive portion joined to a number of negative electrons, which neutralize the positive charge, and that, when the gas is ionized, some of its atoms liberate each one electron, and become themselves positively charged.

Radioactivity

523. Becquerel Rays. — Not long after the discovery of the Roentgen rays the French physicist Becquerel discovered that similar effects to those produced by them could be obtained from masses of uranium or of ores containing uranium. The activity of the uranium was very slight, and several days were required to obtain the photographic effects by which the properties of these rays were studied. They were found to be in general similar to those of the Roentgen rays, but certain differences were noticed, which could not be fully studied on account of the feebleness of the action. The action, however, was considered to be due to rays like the Roentgen rays.

524. Radium and Radioactive Substances. — M. and Mme. Curie found that certain uranium-bearing minerals were more active than metallic uranium. As the result of long-continued labor, they succeeded in isolating from pitchblende, an ore of uranium, a small quantity of a salt of a new element, which they named radium.

This element exhibits properties similar to those of uranium, but to a far greater degree. They are called generally radioactive properties. The most delicate and ready way of detecting radioactivity is by the use of a sensitive electroscope, which will gradually be discharged if the air around it is ionized by the radioactive body. Thorium, among the elements already known, was found also to be radioactive, and the discovery of several other new radioactive elements has been announced.

The properties of radium and thorium have been very carefully studied by Rutherford and Soddy and their associates, and the conclusions to which they have come are confirmed by the independent observations of other investigators. Since the quantity of active material under observation is so small that it would be hopeless to attempt to examine it directly by chemical analysis, reliance has to be placed on the indications of the electroscope, which show at what rate the active material is discharging electricity. If the disintegration hypothesis of Rutherford and Soddy is once accepted, it leads to a rational and consistent explanation of the various observations. The results which have been obtained are too complicated for a full analysis, but certain of them can be given. In the description we shall speak of the substance under examination as radium, with the understanding that essentially what is said of radium may be said of thorium also.

When a salt of radium, such as radium bromide or chloride, is heated or dissolved in water, it gives off a substance which is called the radium emanation. This is a gas of high atomic weight. It can be condensed into the liquid or solid state at the temperature of liquid air. No chemical combinations into which it enters have been discovered. According to Rutherford and Soddy's theory, which will be followed in the subsequent description, the atom of emanation is an atom of radium from which a so-called α-particle has been expelled. The α-particle is a particle which carries a positive charge. The ratio of its charge to its mass, found by studying the deflections of streams of α-particles in magnetic and electric fields, is half that found for the hydrogen atom in electrolysis. The experiments of Ramsay and Soddy have demonstrated that the continued emission of α-particles in a small vessel results in the appearance in the vessel of noticeable quantities of the gas helium. Helium is also found present in most of the radioactive minerals. It is believed that as the α-particles give up or lose their positive charges, they become atoms of helium. Since the

atomic weight of helium is 4, we must suppose that the α-particle, in the condition in which its properties have been observed, carries two positive ionic charges.

The emanation does not remain permanent, but disintegrates in a similar manner by the emission of one α-particle, the remainder becoming radium A. Radium A passes through a change into radium B, with the loss of another α-particle, and radium B changes into radium C, with the loss of a β-particle, which moves so slowly that its detection is difficult. The β-particle behaves in every way like the negative electrons observed in the cathode discharge, and is no doubt a negative electron.

Radium C undergoes a peculiar disintegration, for it not only emits an α-particle, but also a β-particle and γ-rays. The β-particles emitted in this case possess much higher velocities than those of the negative electrons in the cathode discharge, approaching in some cases the velocity of light. The γ-rays are apparently Roentgen rays of a very penetrating type. They presumably originate in connection with the emission of the β-particle.

Subsequent changes of a similar sort result in the successive formation of radium D, E, and F. Radium F is believed to be identical with polonium, which was isolated from pitchblende by M. and Mme. Curie, in the course of the investigation in which they discovered radium. Polonium gives forth α-particles, and is no doubt disintegrating into some other substance. It has been suggested that the final, and so far as our knowledge goes, the stable outcome of this series of changes is the element lead.

By this process of disintegration the original radium is gradually destroyed. Rutherford calculates that the time required to reduce a mass of radium by one-half is 1300 years. This being so, it is difficult to account for the amount of radium present in certain minerals except on the hypothesis that radium itself is a disintegration product of some more stable substance. The evidence points to uranium, in the ores of which radium is always found, as being this substance. Uranium is radioactive also, emitting α- and β-particles and γ-rays, and thus shows that the disintegration postulated for it on this hypothesis is really going on. Its transformation is exceedingly slow. Rutherford calculates that half of a mass of uranium will be transformed in 3.4×10^9 years.

The time required for the transformation of the various intermediate stages between radium itself and radium F, or polonium, is very different in the different cases. The shortest time required

to reduce the product by one-half is 3 minutes, in the case of radium A; the longest, 40 years, in the case of radium D. The corresponding time for radium F is 143 days.

The changes which result in this gradual transformation take place in the individual atoms. In the slow transformations only a relatively small number of atoms break up at once, though the actual number is very large; sufficiently large in the case of radium for the emission to constitute streams of negative electrons and positive α-particles. Most of the products of disintegration remain in the mass, and after a time a condition is attained called the condition of radioactive equilibrium, in which as much of each of the active products is formed by the disintegration of the product before it in the series as is destroyed by disintegration. The radioactive minerals are in this condition of radioactive equilibrium. One of the reasons for believing that radium is a disintegration product of uranium is the proof by Boltwood that the ratio of the radium to the uranium present in the uranium minerals is generally the same.

This view is confirmed by the discovery by Rutherford and Soddy that thorium can be separated by a chemical process into two parts, called thorium and thorium X. The thorium thus obtained is at first inactive, but gradually acquires radioactive properties, and then a new quantity of thorium X can be obtained from it. The thorium X, on the other hand, is at first highly radioactive, emitting α-particles and forming an emanation as radium does. In time the thorium X loses its special activity. It is evident that in this case the two constituents to which the radioactive equilibrium of thorium is due have been artificially separated.

A similar separation into an initially neutral state and a highly radioactive state has been effected with uranium and with another recently discovered radioactive element, actinium.

The disintegration of a radioactive element is accompanied by a loss of energy as the α-particles are expelled from the atoms of which they previously formed a part. This energy is obtained as heat in the mass. A mass of radium is thus always at a higher temperature than the air around it. Curie and Laborde found that the heat emitted by a gram of a radium compound amounts to more than 80 therms an hour.

The emission of the α-particles can be directly perceived by the scintillations which they produce when they fall on a screen of Sidot's blende. When a thin plate of this substance is exposed to

the emissions from radium or polonium and viewed through a magnifier, the field of view is filled with sparkling points of light, each of which is produced by the impact of an α-particle.

Radioactive properties which have been shown to be similar to those of the emanation of radium are exhibited by specimens of water drawn from wells in all parts of the world. The examination of the ordinary rocks has shown that radioactivity is exhibited by them also. Such observations as these show that radioactive bodies are very widely distributed in the earth's crust.

It is an interesting and important question whether all of the ordinary elements, at least those of considerable atomic weight, are undergoing a disintegration similar to that of uranium and thorium. Analogy would lead one to adopt this view, and Campbell has probably detected radioactivity from potassium and sodium. Rutherford points out that the difficulty of detecting the disintegration of most elements may result from the velocity of the α-particles expelled by those elements being too low to effect the ionization of the air through which they pass.

525. Electric Theories of Matter. — Kaufmann has examined by experiment the effect produced by a magnetic field on those rays from radium which correspond to the cathode stream. He finds that under the action of the field the stream spreads out into a gradually widening band. By studying the curvatures of different parts of the stream, he finds that they are consistent with the supposition that the electrons in the stream have no mass. It was shown by J. J. Thomson from Maxwell's theory of electricity that when a charged sphere is moving rapidly through the ether, and is acted on by a force tending to change its motion, it will behave as if its mass were greater than the mass of the sphere, by an amount which depends on the electric charge upon it, and the velocity with which it is moving. This general conclusion, that a moving charge will have an apparent inertia, due to its interaction with the ether, has been confirmed by other investigators, and especially by Abraham, with whose equations Kaufmann interpreted his results. Kaufmann finds that the paths of the electrons sent out by the radium are such as to indicate that the mass of the electron depends upon its velocity, and in such a way as it would if the electron possessed no mass except the apparent mass due to its electric charge. If this result is accepted, we must conclude that the electron is a small portion of electricity, separate from matter. This conclusion has been confirmed by the experiments of Bucherer.

The apparent mass which a moving electric charge can have indicates a theory of matter in which the material atoms are assumed to be assemblages of electrons, moving around one another in very small orbits but with enormous velocity. The ether is assumed to be a continuous medium, possessing a peculiar elastic reaction to torques or twists. Theories resting on this general supposition have been developed by H. A. Lorentz and by Larmor, and have been shown capable of accounting for most, if not all, of the phenomena of light and electricity, in their relations to moving as well as to stationary bodies.

EXERCISES

Principle of Moments. Principle of Work.

1. A weight of 100 lb. is in a wheelbarrow with its center of gravity 2 ft. from the axis of the wheel. The handles are 5 ft. from the same axis; what force must be used to just lift them?

Ans. 40 lb.

2. A nut placed 1 in. from the hinge of a nut-cracker can be broken by a force of 30 lb. applied to the handles at a point 4 in. from the hinge. How heavy a weight will be needed to break the nut, if placed directly on it?

Ans. 120 lb.

3. Two parallel forces of 5 lb. and 9 lb. are acting upwards at the ends of a bar of weight 2 lb. and 36 in. long. (*a*) What additional weight will make equilibrium; (*b*) at what point must it be applied?

Ans. (*a*) 12 lb. (*b*) At a point 24 in. from the 5-lb. weight.

4. Two forces of 6 lb. and 9 lb. are at distances 12 in. and 8 in. from a given line in a plane perpendicular to their plane. Find the distance of their resultant from that line. *Ans.* 9.6 in.

5. Given the parallel forces and their coördinates with respect to x- and y-axes; 5 (3, 4), 8 (2, −3), −2 (7, −5), 13 (−3, 8), find their resultant.

Ans. $R = 24$, $x_r = -\frac{1}{12}$, $y_r = \frac{5}{12}$.

6. Find approximately the center of gravity of a sledge hammer with a head weighing 16 lb. and 4 in. in diameter, and a handle weighing 2 lb. and 2½ ft. long. *Ans.* $\frac{1}{3}$ in. from the center of the head.

7. If the boards in Examples, I, 2, are 2½ ft. long, and the man weighs 150 lb., find what pull he exerts when his feet are 1½ ft. from the hinges.

Ans. $2P = 225$ lb.

8. A man carries a pole of weight W and length l over his shoulder. It rests on the shoulder at a distance d from the end at which it is held by the hand. Find the pull Q exerted by the hand, and the reaction R at the shoulder.

Ans. $Q = W\dfrac{l-2d}{2d}$, $R = \dfrac{Wl}{2d}$.

9. A train of cars rests on a bridge. Show that the reactions at the piers are the same as they would be if the weight of the train was concentrated at its center of gravity.

10. The arms of a balance are 15 cm. long, and the beam weighs 400 gr. A weight of 10 mg. will give a deflection of 4° ($\tan 4° = 0.0699$). Find how far the center of gravity of the beam is below the point of support.

Ans. 0.0536 mm.

11. Given the various weights and their coördinates in the xy plane, 2 (1, 3), 3 (2, 0), 5 (4, 1), 3 (2, −2), 6 (−3, −1), 1 (−2, 3), find the center of gravity.

Ans. $x_r = 0.7$, $y_r = 0.1$.

12. Given the various weights and their coördinates along the rectangular axes of x, y, z, 2 (1, 3, 1), 3 (2, 0, 0), 4 (4, −2, −3), 1 (−6, 1, 3), find the center of gravity. *Ans.* $x_r = 1.8$, $y_r = -0.1$, $z_r = -0.7$.

13. If a clock weight weighing 30 lb. falls 3 ft. in a week, what force P applied to the end of the minute hand, which is 6 inches long, will stop the clock?

Ans. $P = 15/28\pi$ lb.

14. A man of weight W is sitting on a light seat suspended from the movable block of a block and tackle, and keeps himself from falling by holding on to the free end of the rope. If there are 2 pulleys in each block, what force P does he exert? *Ans.* $P = W/5$.

Parallelogram of Forces.

15. Two smooth pins are fixed 4 ft. apart at the same height in the face of a vertical wall. A cord laid over the pins is stretched by weights of 10 lb. hung on its ends. A weight of 16 lb. is hung on the cord midway between the two pins. Find the equilibrium position of the cord.

Ans. Let α represent the angle made by a branch of the cord with the horizontal; then $\sin \alpha = 4/5$.

16. Two forces 3 and 4 have a resultant equal to $\sqrt{37}$; find the angle between the forces. *Ans.* $\alpha = 60$.

17. Three forces of 9, 12, and 15 are in equilibrium; find the angles they make with one another.

Ans. Angle $(9, 12) = 90°$, angle $(9, 15) = 126° 52'+$, angle $(12, 15) = 143° 8'-$

18. A body of weight W is held at rest on a smooth plane, inclined at the angle α, by a string making the angle θ with the face of the plane. (*a*) Find the tension P in the string; (*b*) for what value of θ is it a minimum?

Ans. (*a*) $P = W \sin \alpha / \cos \theta$. (*b*) When $\theta = 0$.

19. A barrel weighing 250 lb. rests on a plane inclined at an angle of 40°. (*a*) What force must be applied to it horizontally to keep it at rest; (*b*) what is the least force that will keep it at rest? *Ans.* (*a*) 209.7 lb. (*b*) 160.7 lb.

20. The ladder of Examples, III, 4, weighs 120 lb. and is 15 ft. long. Its angle of inclination is 24°. Find (*a*) the horizontal push against the cleat, and (*b*) the total reaction at the cleat. *Ans.* (*a*) 26.7 lb. (*b*) 122.9 lb.

21. Find the push against the cleat when the ladder of the last Exercise is inclined 80° from the vertical. *Ans.* 340 lb.

22. The rod in Examples, III, 5, weighs 20 lb. and the angles α and β are each 45°. Find the tensions in the cords and the angle γ made by the upper cord with the vertical.

Ans. Tension in lower cord $\sqrt{50}$ lb., in upper cord $\sqrt{250}$ lb.; $\tan \gamma = \frac{1}{3}$, $\gamma = 18° 26'$.

23. A smooth pin is set in a wall, near a corner, horizontal and parallel with the face of the other wall, and a smooth rod rests across the pin with one end resting against the wall. The distance of the pin from the wall is $1/n$th the length of the rod. Show that the angle α made by the rod with the normal to the wall is given by $\cos^2 \alpha = 2/n$.

24. Three smooth pins, projecting horizontally from a vertical wall, are set at the corners of an equilateral triangle with its base horizontal, and a cord hanging over them supports two equal weights P attached to its ends. Find the reactions at the pins.

Ans. At the upper pin, $2P \cos 30° = P\sqrt{3}$; at each of the lower pins, $2P \sin 15° = P \dfrac{\sqrt{3} - 1}{\sqrt{2}}$.

Static Friction.

25. A block of wood will just start sliding on a wooden inclined plane of which the inclination is 24°. Find the coefficient of friction. *Ans.* 0.445.

26. Weights $W = 10$ lb. and P unknown are arranged as in Examples, IV, 2. The weight W slips down the plane when $\alpha_1 = 80°$, and up the plane when $\alpha_2 = 15°$. Find the coefficient of friction, and the value of P.

Ans. $\mu = 0.637$, $P = 8.74$ lb.

27. If the coefficient of friction of the ladder with the floor in Examples, IV, 4, is 0.45, and the angle of inclination is 40°, to what fraction of the length of the ladder can the man ascend before slipping takes place? *Ans.* $d/l = 45/83.9$.

28. A weight of 200 lb. rests on a rough floor, with which its coefficient of friction is 0.8. A force of 140 lb. is available to slide it along the floor. At what angle α from the vertical will the force begin to move the weight? (See Examples, IV, 5.) *Ans.* $\alpha = 24° 30'$ about.

29. A load of 200 lb. is in a wheelbarrow, with its center of gravity 2 ft. from the axle of the wheel. The handles are 5 ft. from the axle. If the coefficient of rolling friction is 0.15 (*a*), what is the angle with the vertical made by the forces exerted by the hands when the barrow starts moving; (*b*) what is the force exerted by the hands? *Ans.* (*a*) $12° 40'$, about. (*b*) 82 lb.

Velocity and Acceleration.

30. A duck flying with the speed of 50 miles an hour passes a gunner 60 yds. away. If the gunner shoots so that the shot travel at right angles to the path of the duck, and if the shot have a speed of 800 ft. per second, how far ahead of the duck must the gun be aimed in order to hit? *Ans.* 16.5 ft.

31. An army 10 miles long marches 20 miles in 10 hours. A courier, starting from the head of the army at the beginning of the march, goes to the rear and returns to the head just as the march is ended. How far does he travel? *Ans.* 32.36 miles.

32. A man of height h walks with a speed of v feet per second on a level street away from an electric light hung at a height of b ft. above the ground. What is the speed of the end of his shadow? *Ans.* $bv/(b-h)$.

33. A fleet is traveling at the rate of 10 knots (a knot is a speed of 1 nautical mile per hour). The last ship has dropped behind its station 6 miles in 3 hours.

(*a*) At what rate has it been traveling?

(*b*) How fast must it travel to resume its place in 4 hours?

(*c*) If it reaches its station in 4 hours by increasing its speed uniformly what must be its acceleration?

(*d*) In this case what speed will it have when it reaches its station?

(*e*) What negative acceleration must it exceed from the instant at which it reaches its station to prevent its reaching the next ship ahead, the ships being spaced ¼ mile apart?

Ans. (*a*) 8 knots; (*b*) 11.5 knots; (*c*) ¼ knots / hours; (*d*) 15 knots; (*e*) 50 knots / hours.

34. Show that the distances passed through by a falling body in successive equal intervals of time are proportional to the successive odd numbers.

35. Show that if we start measuring distance from a point in the path of a falling body through which it passes at a particular instant, and start measuring time from that instant, the differences of the successive distances passed over in successive equal intervals of time are equal to gt^2, where t is the time interval chosen.

36. Find the straight line of quickest descent from a point to a given straight line.

37. Find the times at which a body projected vertically upward with an initial velocity v will pass through a given point at the height s.

$$Ans.\ t = \frac{v \mp \sqrt{v^2 - 2gs}}{g}.$$

38. A bullet is projected from the ground with a speed of 96 ft. per second at an angle of 30° with the horizontal. Find (*a*) the height to which it rises; (*b*) the time of flight; (*c*) the range, taking $g = 32$.

Ans. (*a*) 36 ft.; (*b*) 3 sec.; (*c*) $144\sqrt{3}$ ft.

39. The same bullet is similarly projected from a platform 28 ft. above the ground. Find (*a*) the time of flight; (*b*) the range.

Ans. (*a*) 3½ sec.; (*b*) $168\sqrt{3}$ ft.

40. The velocity of the bullet from a small rifle is 300 meters per second. What must be the actual direction of projection in order that the bullet shall hit a mark at the level of the muzzle and 200 meters away? Use $g = 9.8$ meters/sec^2.

Ans. 37′.5, about, above the horizontal.

41. What is the period of a simple pendulum 18 ft. long, taking $g = 32.2$ ft./sec^2.?

Ans. 2.35 − sec.

42. What is the length in centimeters of the seconds pendulum on the moon, where the acceleration of gravity is ⅙ that on the earth? *Ans.* 16.55 cm.

43. A simple pendulum 99.093 cm. long beats seconds at the equator. What is the value of g? *Ans.* 978.01 cm./sec^2.

44. The sun is 928×10^5 miles from the earth, and the length of the year is 365¼ days. (*a*) What is the average velocity of the earth in its orbit, taken as circular; and (*b*) what is its acceleration toward the sun?

Ans. (*a*) 18.47 miles/sec. (*b*) 0.2328 inch/sec^2.

45. The earth's equatorial radius is approximately 63.78×10^5 meters; what is the amount by which gravity at the equator ($g = 978$) would be increased if the earth did not rotate? *Ans.* 3.37 cm./sec^2.

46. How fast would the earth have to rotate in order to reduce gravity at the equator to zero? *Ans.* 17.057 times as fast as at present.

Mass and Momentum.

47. The masses used with the Atwood's machine weigh 200 gr. each. The overweight weighs 10 gr. (*a*) How far will the system move from rest in 3 seconds? (*b*) What is the tension in the cord?

Ans. (*a*) 107+ cm. (*b*) 2.008×10^6 dynes = 204.9− grams-weight.

48. What will be the velocity at the end of 4 seconds of the system of the Atwood's machine, if the masses weigh 200 gr. each and the overweight 15 gr.?

Ans. 141+ cm./sec.

49. A force of 20 grams weight carries the system of an Atwood's machine over 160 cm. in 4 seconds. What is the mass of the system (*a*) in gravitation units, (*b*) in absolute units? *Ans.* (*a*) 1; (*b*) 980 grams.

50. A weight of 2 kg. will stretch a spring to a certain length. If the spring is attached to a mass of 6 kg. lying on a smooth table and is stretched horizontally to that length, before the mass is released, with what acceleration will the mass start moving? *Ans.* 326⅔ cm./sec.²

51. A baseball weighing 9 oz. is pitched with a velocity of 90 ft./sec., and when struck by a bat moves in the opposite direction with a velocity of 120 ft./sec. (*a*) What is the impulse? (*b*) If the ball is in contact with the bat for $\frac{1}{10}$th of a second, what is the average force exerted on it?

Ans. (*a*) 118 poundal-seconds. (*b*) 1180 poundals.

52. A bullet weighing 2 gr. is shot from a rifle barrel 50 cm. long and striking against a suspended wooden block weighing 1998 gr., in which it embeds itself, swings it out so that its center of gravity rises $\frac{1}{16}\frac{5}{6}$ cm. Assuming that the gases in the barrel exert a constant force on the bullet, find the value of that force.

Ans. 50×10^6 dynes.

Work and Energy.

53. Compare the kinetic energy of the bullet of Ex. 52 with that of the block and bullet after the impact.

Ans. K of bullet $= 5 \times 10^6$ ergs, K of block $= 5 \times 10^6$ ergs.

54. What would be the kinetic energy of a raindrop weighing 0.25 gram falling to earth from a cloud 1000 meters high, if it were not for the resistance of the air?

Ans. 245×10^6 ergs.

55. If 2.5 centimeters of rain fall from clouds 1000 meters high, how much energy in kilogram-meters would be expended by friction in the atmosphere by the amount which falls on a square meter, if it reaches the ground with a velocity of 30 meters a second? *Ans.* 23,952 kilogram-meters.

56. What constant force, in pound weights, will lift 10 lb. from rest over 25 ft. in 5 seconds? *Ans.* 10.62+ lb.

57. Show that the energy of a railway train traveling with the velocity of 30 miles an hour would be sufficient to lift the train 30.25 ft.

Collisions.

58. Two ivory balls of equal mass m, for which the coefficient of restitution is 0.85, are moving in the same direction, the one in front with the velocity $u_1 = 100$ cm./sec., the one in the rear with the velocity $u_2 = 150$ cm./sec. Find (*a*) the velocities after collision, (*b*) the loss of kinetic energy by collision.

Ans. (*a*) $v_1 = 146.25$ cm./sec., $v_2 = 103.75$ cm./sec.; (*b*) $173.2\,m$ ergs.

59. If the same ivory balls as in the last Exercise are moving in opposite directions with the same speeds as before, find (*a*) the velocities after collision; (*b*) the loss of kinetic energy by collision.

Ans. (*a*) $v_1 = -131.25$ cm./sec., $v_2 = 81.25$ cm./sec.; (*b*) $4336\,m$ ergs.

60. Investigate the results of the collision of two steel balls, one weighing 99 kg., the other 1 kg., moving with velocities of 10 meter/sec. and −40 meter/sec. respectively. The coefficient of restitution is 0.7.

Ans. $v = 9.5$ m./sec., $v_1 = 9.15$ m./sec., $v_2 = 44.15$ m./sec., kinetic energy lost 631+ kg.m².sec⁻².

61. If a sphere is dropped from a height h upon a heavy slab, show that the height k_n to which it will rise after n collisions is given by the formula $k_n = he^{2n}$.

62. If the sphere of Ex. 70 is dropped from the height of a meter, how many times will it rebound from the slab to the height of over a millimeter; the coefficient of restitution being 0.7? *Ans.* 9 times.

63. A wheel 3 ft. in diameter rolls down an inclined plane with a linear acceleration of 12 ft./sec.^{2} What is (*a*) its angular acceleration around its axis; (*b*) what is its angular velocity at the end of 4 seconds? *Ans.* (*a*) $\alpha = 8$ rad./sec.^{2}, (*b*) $\omega = 32$ rad./sec.

Angular Motion.

64. Show that the centripetal force on a mass m moving in a circle of radius r with the velocity v may be expressed by $m\omega^2 r$.

65. A heavy disk weighing 50 lb. and 4 ft. in diameter is set in rotation by a cord drawn off its axle, 6 in. in diameter, with a force of 20 lb. weight. Find its angular acceleration. *Ans.* $\alpha = 1.6$ rad./sec.^{2}

66. A cylinder 2 ft. in diameter and weighing 200 lb. is set in rotation by a weight of 50 lb. hung to the end of a cord wrapped round the cylinder. Find (*a*) the angular acceleration, (*b*) the velocity of the weight after 3 seconds; (*c*) and show that the kinetic energy of the system after 3 seconds is the same as that which the weight would have had if it had fallen freely through the same distance.

Ans. (*a*) $\alpha = \frac{8}{3}$ rad./sec.^{2}; (*b*) $\omega = 32$ rad./sec.; (*c*) $K = 75 \times 32^2$ lb.ft.^{2}/sec.^{2}

67. An impulse equal to a force of 5000 dynes applied for one second is applied tangentially to the rim of a flat disk weighing 200 gr. and with a diameter of 6 cm. Find the resulting motion.

Ans. A velocity of the center of the disk equal to 25 cm./sec. and an angular velocity of $\frac{50}{3}$ rad./sec.

68. Find the center of percussion of the disk in the last Exercise.

Ans. $1\frac{1}{2}$ cm. from the center.

69. Two small heavy masses, each of 500 gr., are fastened to the ends of a stiff rod 40 cm. long, whose mass may be neglected. An impulse of 10^6 dynes is applied to one of the masses, at right angles to the rod. Find (*a*) the resulting velocities of the center of mass and of each mass, (*b*) the kinetic energy imparted to the system.

Ans. (*a*) The mass to which the impulse is applied has a velocity of 200 cm./ sec., the other one is instantaneously at rest; (*b*) 10^7 ergs.

70. Prove that, if the moment of impulse Jp applied to a body free to rotate about an axis changes its angular velocity from ω_0 to ω, the kinetic energy given to the body is $Jp(\omega + \omega_0)/2$.

71. A rectangular bar 100 cm. long and 2 cm. broad is swung from a knife-edge set 10 cm. from one end. Find (*a*) where another knife-edge shall be set so as to construct a reversible pendulum, assuming that the weights of the knife-edges may be disregarded; and (*b*) the period of vibration, if $g = 980$.

Ans. (*a*) 60.8 cm.; (*b*) 0.768 sec.

72. A sphere of platinum 6.056 cm. in diameter is suspended by a weightless thread from a point of support and swings as a pendulum, with a period of $\frac{1}{2}$ a second. Find the distance from the point of support to the center of the sphere.

Ans. $R = 24.7$ cm. when $g = 980$.

73. If a wooden sphere of the same weight as that of the platinum sphere of the last Exercise, but of diameter 22.854 cm., is suspended so that the distance from its center to the point of support is the same as that of the platinum sphere, find its period when it is swung as a pendulum. *Ans.* 0.519 sec.

74. Where should a thin hoop be struck so that it shall roll without slipping on a smooth floor? *Ans.* At the highest point.

75. Show that the acceleration of a thin hoop or cylindrical tube, rolling down an inclined plane, is half that of a body sliding without friction down the same plane.

76. If two weights of 100 gr. and 150 gr. are hung on the ends of a cord that passes over a grooved pulley which is held fast so that the cord slides in the groove, and if the heavier weight falls 60 cm. in the first second, find the tensions in the cords and the coefficient of friction.

Sliding Friction.

Ans. $T_1 = 129{,}000$ dynes, $T_2 = 110{,}000$ dynes, $\mu = 19/239$.

77. If a bicyclist starting with a velocity of 13.2 ft./sec. coasts on a level road and comes to rest at the end of 522 ft., find the coefficient of friction.

Ans. $\frac{1}{66}$.

78. A lead pipe weighing 15 lb. and 2 ft. in diameter is put in place of the tire of a bicycle wheel. The circle of the ball bearing is 0.6 in. in diameter. The wheel is started rotating twice a second, and comes to rest in 5 minutes. Find the coefficient of friction. *Ans.* $2\pi/75$.

Hydrostatics.

79. A block of aluminium, weighing 75 gr. in air, weighs 45 gr. in water. Find its relative density. *Ans.* 2.5.

80. If the breaking weight of a cord is 60 lb., how great a mass of copper can the cord sustain if the copper is immersed in water? *Ans.* 67.6 lb.

81. If a cylindrical column of water 40 cm. high weighs 480 gr., how much will a column of the same height weigh if 10 cm. of it are mercury? *Ans.* 2091.5 gr.

82. What will be the apparent weight of a crystal of quartz, weighing in air 250 gr., if it is immersed in ethyl alcohol? *Ans.* 175.4 gr.

83. Find approximately the weight in pounds of a cubic foot of water. *Ans.* 62.5 lb.

84. What corrections must be applied to get the true weights in vacuum in weighing quartz with brass weights in air?

Ans. The weight indicated by the brass weights must be multiplied by 1.000336, approximately, to get the true weight of the quartz.

85. If the relative density of gold is 19.3, and that of silver 10.5, what was the relative density of the crown in Archimedes' problem, if half the weight was silver? *Ans.* 13.6.

86. Using the result of the last Exercise, show that in Archimedes' method (p. 135) of solving the problem of the crown, the difference in weight of the gold and the silver in water was 0.0434 times the weight of the gold, and that the difference of weight of the gold and the crown was half that.

87. Find the height of the mercury column which will exert a pressure of 10^6 dynes/cm.^{2} at a place where $g = 980$. *Ans.* 75.05 cm.

88. Find the pressure in dynes per square centimeter of a column of glycerine 3 meters high. *Ans.* 3.705×10^5.

89. Find the pressure of 1 atmosphere (*a*) in gram weights per square centimeter, (*b*) in pounds per square inch. *Ans.* (*a*) 1033; (*b*) 14.7.

90. A balloon weighs 2 gr. (*a*) What must be its volume when filled with hydrogen in order that it will just float in air; (*b*) what is the mass of hydrogen contained in it? *Ans.* (*a*) 1662.5 c.c.; (*b*) 0.1496 gr.

91. What will be the apparent weight of the same balloon when in an atmosphere of carbon dioxide? *Ans.* Negative and equal to 1.132 gr. weight.

Gravitation.

92. Show from Huygens's theorem of centrifugal force that if different bodies move uniformly in circles about a common center, in such a way that the squares of their periodic times are proportional to the cubes of the radii of their orbits, the accelerations of the bodies are inversely as the squares of the radii of their orbits.

93. From Equation 71 show that if M represents the mass of the sun, P the mass of a planet, T and a the periodic time and distance of the planet, T_1 and a_1 the periodic time and distance of a satellite of the planet, the ratio of the mass of the sun to the mass of the planet is given by $\frac{M}{P} = \frac{T_1^2 a^3}{T^2 a_1^3}$.

94. The mean distance of Jupiter from the sun is 483×10^6 miles, that of its outermost satellite from Jupiter is 1.169×10^6 miles. The period of Jupiter is 11.86 years, that of the satellite 16 days 16½ hours. Treating the orbits as circular, find the approximate ratio of the mass of the sun to the mass of Jupiter. *Ans.* 1046.

95. Find the periodic time of a small imaginary satellite describing a circular orbit around the earth, just above its surface at the equator; it being given that the earth's equatorial radius is 6,378,301 meters; and that g at the equator equals 9.814 meter/sec.^{2} when the effect of the earth's rotation is allowed for. *Ans.* 1.407 hours.

96. The mass of the moon is $\frac{1}{80}$ that of the earth; its mean distance is 60.27 times the earth's equatorial radius; its periodic time is 27,322 days. Compare its motion with that of the imaginary satellite of Exercise 107 to show that they conform to Kepler's third law when modified to take account of the mass of the moon. (Examples, XI, 3.)

97. What is the gravitational force between two balls of lead, each weighing 500 kg., when their centers are 100 cm. apart? *Ans.* 1.625 dynes.

98. What is the acceleration of gravity at the top of a mountain 3 miles high, in latitude 50°, where $g = 981$? *Ans.* $F = 979.53$.

Elasticity.

99. What will be the proportionate diminution of the volume of a mass of water on an increase of pressure of 1 atmosphere? *Ans.* 0.00005.

100. What will be the elongation of a steel wire 3 meters long and with a cross section of 2 sq. mm., when stretched by a weight of 12 kg., if Young's modulus is 2×10^9 in grams per square centimeter? *Ans.* 0.9 mm.

101. If a brass wire 4 meters long and with a cross-section of 3 sq. mm., when stretched by a weight of 9 kg., is elongated by 1.5 mm., what is Young's modulus for the wire? *Ans.* 12×10^8 in grams per square centimeter.

102. If a cylindrical brass wire 50 cm. long and with a radius of 1 mm. is twisted by a couple equal to 200 in grams-weight and centimeters, what will be the amount of torsion, if the modulus of rigidity is 400×10^6 in grams per square centimeter? *Ans.* $\phi = 1/2\pi$.

103. From Examples, XII, 4, and Equation 75 show that the rigidity may be calculated from observations of a body oscillating at the end of a suspended wire by the formula $n = \frac{8\pi lI}{T^2 r^4}$, in which r is the radius of the wire.

104. A mass with a moment of inertia of 490 is hung from the end of a steel wire 100 cm. long, with a radius of 0.5 mm. and oscillates with a period of 0.5 seconds. Find the rigidity of the steel in dynes per square centimeter.

Ans. $n = 980 \times 804.2 \times 10^6$ dynes/cm.^{2}

Capillarity.

105. If the surface tension of a film of soapy water is 50 dynes/cm., what will be the excess of pressure inside a soap bubble 8 cm. in diameter, due to the tension in the film?

Ans. 50 dynes/cm.^{2}

106. If the movable side of the frame described in § 169 is a thin wire 6 cm. long, what must be its weight to just balance the upward tension of the film of soapy water ($T = 50$ dynes/cm.) when the frame is inclined at an angle of 30°?

Ans. 1.22 gr.

107. If we assume the contact angle of water with glass to be zero, what will be the height of the column of water in a vertical capillary tube, of which the diameter is 1.25 mm.? *Ans.* 2.38 cm.

108. We assume that the contact angle of alcohol with glass is zero. A column of alcohol rises in a tube 0.72 mm. in diameter to the height of 1.54 cm. The specific gravity of the alcohol is 0.79. Find the surface tension.

Ans. 21.5 dynes/cm.

109. The contact angle of mercury with glass is about 135°. The top of a column of mercury sinks in a tube 0.4127 mm. in diameter to a depth of 1.47 cm. below the external surface. Find the surface tension. *Ans.* 466.7 dynes/cm.

Gases.

110. A mass of air under atmospheric pressure occupies 100 c.c. The pressure on it is increased by 400 grams-weight per square centimeter. Find its volume. *Ans.* 72.09 c.c.

111. Air is confined in a vertical tube, the lower open end of which is plunged in a deep vessel of mercury. When the mercury surface in the tube is level with that in the vessel it is 60 cm. distant from the upper end of the tube. The tube is then lifted until the top of the mercury column in the tube is 26 cm. above the free surface of mercury in the vessel. How far will it be from the upper end of the tube? *Ans.* 91.2 cm.

112. A rubber balloon 32 cm. in diameter and weighing 2 gr. is filled with hydrogen. The tension in the rubber membrane is 4.052×10^6 dynes per centimeter. What volume would the gas occupy if it were run into a gasometer at atmospheric pressure? What is the lifting force of the balloon?

Ans. 25,736 c.c., 16.71 grams-weight.

113. What is the average kinetic energy of a molecule of hydrogen at 0° C. and under atmospheric pressure? What is the whole translational kinetic energy of a cubic centimeter of hydrogen?

Ans. 3.73×10^{-14} ergs. 1.5×10^6 ergs.

Sound.

114. Calculate the velocity of sound in an atmosphere of hydrogen at 0° C. (The elasticity is 1.41 times the pressure.) *Ans.* 1.26×10^5 cm./sec.

115. From the result of the last Exercise calculate the velocity of sound in an atmosphere of carbon dioxide at 0° C. *Ans.* 2.7×10^4 cm./sec.

116. A string is stretched with a weight of 1000 gr. A meter's length of the string weighs 1.5 gr. Find the velocity of a wave on the string. *Ans.* 80.8 m./sec.

117. A string, a meter's length of which weighs 1.5 gr., and which is stretched by a weight of 250 gr., is set in vibration by a vibrating rod which makes 14 double vibrations in a second, and stationary waves are established on it. Find the distance between the nodes. *Ans.* 1.44 meters.

118. A thread whose mass per meter is 0.15 gr. is stretched by a weight of 50 gr. and set vibrating in stationary waves by a tuning fork whose frequency is 60. Find the distance between the nodes. *Ans.* 47.6 cm.

119. A wire 1 meter long, whose mass per meter is 0.5 gr., is stretched by a weight of 2 kg. Find the frequency of the fundamental vibration. *Ans.* 99.

120. What is the frequency of the fundamental vibration in an open pipe 60 cm. long? *Ans.* 276.

121. What is the length of the open pipe in which the frequency of the fundamental vibration is 426⅔? *Ans.* 38.85 cm.

122. What is the frequency of the vibrations in an open pipe 16 ft. long? *Ans.* 34, nearly.

123. Find the frequencies of the fundamental and the first two harmonics in a stopped pipe 2 ft. long. *Ans.* 136, 408, 680.

124. Two pipes tuned to *si* and the upper *do* of the major scale make 12 beats a second when sounded together. What are their frequencies? *Ans.* 180 and 192.

Heat.

125. Express (*a*) 72° Fahr. on the Centigrade scale; (*b*) −40° Fahr. on the Centigrade scale; (*c*) −273° C. on the Fahrenheit scale. *Ans.* (*a*) 22.2° C., (*b*) −40° C., (*c*) −459.4° C.

126. How much heat in calories will raise the temperature of 1 lb. of water 1° Fahr.? *Ans.* 0.252 calorie.

127. A mass of water at 100° is mixed with a mass of mercury, having equal volume, at 10°. What is the resulting temperature? *Ans.* 72.1°.

128. What must be the ratio between the volumes of two masses of mercury and water, which will produce by their mixture a resultant temperature equal to the mean of their original temperatures? *Ans.* 2.2.

129. A copper calorimeter weighing 200 gr. contains 500 gr. of water at 20°. How much water at 90° must be poured into it, to bring the final temperature to 55°? *Ans.* 518 gr.

130. What will be the resultant temperature if 500 gr. of water at 90° are poured into a copper calorimeter weighing 200 gr., and containing 500 gr. of water at 20°? *Ans.* 54.3°.

131. A block of silver weighing 1000 gr. at 95° is immersed in 500 gr. of water at 20° contained in a silver calorimeter weighing 200 gr. The resultant temperature is 27.4°. What is the specific heat of silver? *Ans.* 0.056.

132. Equal quantities of heat are supplied by an electric current to liquids contained in two similar calorimeters. Show that the specific heats of the liquids are in the inverse ratio of the products of their masses and changes of temperature.

133. In the arrangement described in Ex. 132, one of the calorimeters contains 700 gr. of water, and its temperature rises 10°; the other contains 735 gr. of a solution of milk sugar, and its temperature rises 10.019°. Find the specific heat of the solution. *Ans.* 0.9506.

134. A mass of 200 gr. of ice is melted in 500 gr. of water at 90°, contained in a silver calorimeter weighing 200 gr.; the resultant temperature is 42.2°. Find the latent heat of ice. *Ans.* 79.97 therms.

135. Using the same calorimeter as in Ex. 134 and 500 gr. of water at 90°, find the resultant temperature when 511.2 gr. of ice are melted in it, taking the latent heat of ice to be 80. *Ans.* 5°.

136. The specific heat of fluid acetic acid is 0.5, its melting point is 5°. 400 gr. solid acetic acid at 5° are melted in 500 gr. of acetic acid at 90° contained in a silver calorimeter weighing 200 gr.; the resultant temperature is 14.97°. Find the latent heat of acetic acid. *Ans.* 44.4 therms.

137. When 15 gr. of steam are condensed in 250 gr. of water at 20° C. the resulting temperature is 54.92° C. Find the latent heat of steam. *Ans.* 536.9 therms.

138. If 40 gr. of steam are condensed in 400 gr. of water at 0° C. what is the resulting temperature? *Ans.* 57.9° C.

139. How much does the length of an iron rail 40 ft. long at 0° C. change by reason of a change of temperature from −20° C. to 35° C. *Ans.* 0.28 inch.

140. A clock has a pendulum suspended by an iron rod, and is adjusted at 18° C. to beat seconds. How will its rate change if the temperature rises to 35° C.? *Ans.* It loses about 7.8 seconds a day.

141. If aluminium and iron rods 80 cm. long at 10° C. are clamped together at one end, and if the temperature is raised to 100° C., how much will the aluminium protrude beyond the iron? *Ans.* 0.994 mm.

142. The dimensions of a rectangular block of vulcanite are 25 cm., 10 cm. and 4 cm. at 0° C. How much will its volume increase when its temperature is raised from 20° C. to 80° C.? *Ans.* 11.45 c.c.

143. If the pressure of a gas in a fixed volume is 76 cm. of mercury at 15° C., and then by reason of a rise of temperature becomes 114 cm. of mercury, what is the new temperature? *Ans.* 159° C.

144. If the temperature of a gas is raised from 27° C. to 100° C. and the pressure is doubled, what is the ratio of the resulting volume to the original volume? *Ans.* 0.62.

145. What is the value for air of R in Equation 100 when the mass of air is one gram, the pressure is 1.013×10^6 dynes per square centimeter (1 atmosphere) and the temperature is 0° C. *Ans.* 2.87×10^6.

146. What is the value for hydrogen of R in Equation 100 under conditions similar to those of Ex. 145. *Ans.* 41.23×10^6.

147. If masses of different gases are taken which are equal in grams to the numerical value of the molecular weights of the gases, and the conditions are the same as those of Ex. 145, show that the value of R will be the same for all of them and will be equal to 82.46×10^6.

148. Compare the result of Ex. 147 with the value of R calculated from the Equation of § 282 and the data given in Table VII.

149. The measured volume of hydrogen collected in a graduated tube over water is 205 c.c., the pressure is 740 mm. of mercury, the temperature 20° C. Find the volume of the dry hydrogen at atmospheric pressure and 0° C. *Ans.* 181.6 c.c.

150. In Rumford's experiment the amount of heat developed in heating the metal and the water was estimated by him as sufficient to raise the temperature of 26.58 lb. of water 180° Fahr. The time of the experiment was 2 hours 30 minutes. The work was actually done by two horses, but Rumford says that it could easily have been done by one horse. Calculate from these data a value of the mechanical equivalent of heat, assuming that two-thirds of a horse-power was expended in producing heat. *Ans.* 690 ft.-lb. per lb. Fahr. degree.

151. The work done by a gas which expands when its temperature rises one degree at constant pressure is equivalent to the heat absorbed by the expanding gas in excess of that used in raising its temperature, i.e. to the difference between the heat capacities of the gas under constant pressure and at constant volume. The specific heat of air under constant pressure is 0.23788; that at constant volume is 0.16930. Calculate the mechanical equivalent of heat. (This is the method used by Mayer.) *Ans.* 41.84×10^6 ergs.

152. An iron meteorite weighing 2 gr. enters the atmosphere with a velocity of 8 kilometers per second. How much heat is developed by its friction with the atmosphere before it burns up. Show that if one-tenth the heat developed were retained in the meteorite its temperature would rise far above the boiling point of iron. *Ans.* 1.5×10^4 therms.

153. How much heat is developed when a loaded timber sled weighing 800 lb. slides down a hill 400 ft. long at an angle of 20°, at a uniform rate, taking the mechanical equivalent as 772 ft.-lb. per lb. Fahr. degree?

Ans. About 142 lb. Fahr. degrees.

154. What is the efficiency of a Carnot's engine working between the temperatures 200° C. and 40° C.? *Ans.* 0.338.

155. What is the gain of entropy when 10^8 ergs of heat pass by radiation from boiling water to melting ice? *Ans.* 98,200 units of entropy.

Mirrors.

156. The radius of a concave mirror is 10 cm. A point source is placed 40 cm. away from the mirror. Find the position of the focus. *Ans.* 5.71 cm. from the mirror.

157. The source and focus are respectively 60 cm. and 6 cm. from a concave mirror. Find its radius. *Ans.* 10.9 cm.

158. A source is at first 10 meters from a convex mirror, and is then moved up to a distance of 2 meters from the mirror; how much does the focus move if the radius of the mirror is 40 cm.? *Ans.* 1.426 cm.

159. Find a formula for the radius r of a spherical mirror, when the heights H and h of object and image are given, and the distance d between them.

$$Ans.\ r = \frac{2dHh}{H^2 - h^2}.$$

160. An object 80 cm. high is distant 10 meters from a concave mirror whose radius is 50 cm. Find the height of the image. *Ans.* 2 cm.

161. Show that the images of all infinitely distant sources, the rays from which do not make a large angle with the axis, will lie on a plane perpendicular to the axis through the principal focus (the focal plane).

162. The sun subtends an angle of about 30′. What will be the diameter of its image formed by a concave mirror of 40 ft. focal length?

Ans. 4.2 inches, nearly.

Refraction.

163. Show that the index of refraction n (wg) between water and glass is equal to the ratio of the index n (ag) between air and glass, and the index n (aw) between air and water. To do this use the fact that if plane parallel plates of different media are superposed, a ray falling from air on the first plate will emerge into air from the last plate in a direction parallel with its original direction.

164. Find the critical angle (a) for water, (b) for diamond.

Ans. (a) 48° 35′; (b) 23° 53′.

165. A "right-angled prism" of crown glass has its two faces at right angles to each other and the base making equal angles with the faces. Show that a beam of light falling perpendicularly on one of the faces is totally reflected at the base, and emerges with undiminished intensity from the other face in a direction perpendicular to its original direction.

166. Prove that the deviation of a ray by a prism is equal to the sum of the angles of incidence and emergence diminished by the refracting angle of the prism.

167. Find a formula by which the index of refraction of the material of a prism can be calculated from an observation of the minimum deviation and the refracting angle.

$$Ans.\ n = \frac{\sin\frac{\delta+\theta}{2}}{\sin\frac{\theta}{2}}.$$

168. A point source is in water at a point at a distance s below the surface. Find a formula for the point at which a ray emerging from the surface at a point near the point where the vertical ray from the source emerges from the surface will intersect the vertical ray. All such rays will seem to come from a point at a distance d below the surface, which will be a virtual focus.

Ans. $d = s/n$ for all rays for which the angles of incidence and refraction are so small that the ratio of their sines may be set equal to the ratio of their tangents.

169. A ray falls at an angle of 30° upon a plane parallel plate of crown glass 10 cm. thick. Find the sideways displacement of the ray on emergence.

Ans. 1.98 cm. nearly.

170. How much will the apparent distance of an object be diminished if it is viewed through a plane parallel plate of flint glass 10 cm. thick?

Ans. 3.8 cm. nearly.

Refraction at Spherical Surface.

171. Locate the focus of a source at a distance of 40 cm. from a concave spherical surface separating air from crown glass, and with radius $r = 10$ cm. *Ans.* $u' = 20$ cm. nearly.

172. Find the focal lengths of the surface described in the last Exercise.

Ans. $U = -19.23$ cm., $U' = 29.23$ cm.

173. Locate the foci of a source at the distances (*a*) of 50 cm. and (*b*) of 25 cm. from a convex spherical surface separating air from a glass of which the index of refraction is 1.5, when the radius of the surface is $r = -20$ cm.

Ans. (*a*) $u' = -300$ cm.; (*b*) $u' = 100$ cm.

174. If a source is at the distances (*a*) $x = -10$ and (*b*) $x = 15$ from the first principal focus of the surface described in the two preceding Exercises, find the distances x' of its focus from the second principal focus.

Ans. (*a*) $x' = 240$; (*b*) $x' = -160$.

175. An object 30 cm. high is placed 70 cm. away from a concave spherical surface of radius $r = 10$ cm. separating air from glass ($n = 1.5$). Find the height of its image. This is the case represented in Fig. 173 b. *Ans.* $H' = \frac{3}{8}^0$ cm.

176. An object 20 cm. high is placed at the distance of 50 cm. from a convex spherical surface of radius $r = -20$ cm. separating air from glass ($n = 1.5$). Find the height of its image. *Ans.* $H' = -80$ cm.

177. An object 20 cm. high is situated in a mass of glass ($n = 1.5$) and the rays from it are refracted at a concave spherical surface of radius $s = 10$ cm. separating the glass from air. The object is 60 cm. distant from the surface. Determine (*a*) the position and (*b*) the height of its image.

Ans. (*a*) $v = -40$ cm.; (*b*) $h = -20$ cm.

Thin Lenses.

178. Find the focal length of a convex meniscus of glass ($n = 1.5$), of which the radii are $r = 8$ cm., $s = 5$ cm., and determine the position and height of the image of an object 20 cm. high and 50 cm. distant from the lens.

Ans. $F = \frac{8}{3}^0$ cm., $v = -57.14$ cm., $h = -22.85$ cm.

179. Find the focal length of a double convex lens of glass ($n = 1.5$), of which the radii are $r = -6$ cm., $s = 6$ cm., and determine the position and height of the image of an object 20 cm. high and 106 cm. distant from the lens.

Ans. $F = 6$ cm., $v = -6.36$ cm., $h = -1.2$ cm.

180. The distances of a source and its focus from a convex lens are respectively 40 cm. and -8 cm. Find the focal length of the lens.

Ans. $F = \frac{7}{3}^0$ cm.

181. Find the focal length of a concave lens of glass ($n = 1.5$), of which the radii are $r = 4$ cm., $s = -4$ cm., and determine the position and height of the image of an object 100 cm. high and distant 400 cm. from the lens.

Ans. $F = -4$ cm., $v = 3.96$ cm., $h = 0.99$ cm.

182. Show that, when the real image formed by a convex lens is of the same height as the object, the distance between object and image is four times the focal length.

183. A convex lens is placed so as to throw the image of an object on a screen. The lens is then moved until another image of the same object is thrown on the screen. The distance between the object and screen is a, the distance through which the lens is moved is b. Find a formula by which the focal length of the lens can be determined from this observation.

Ans. $F = \dfrac{a^2 - b^2}{4a}$.

184. The standard distance of an object from a kodak lens of 4 inches focal length is about 15 ft. Find the distances between the image of an object at this standard distance and the images of objects about 5 ft., 10 ft., 20 ft., 50 ft., 1000 ft., away from the lens. (Examples, XVII, 6. Take the various given distances for x.)

Ans. For 5 ft., $+1/67.5$ ft.; 10 ft., $+\frac{1}{17}^0$ ft.; 20 ft., $-\frac{1}{15}^0$ ft.; 50 ft., $-\frac{1}{13}^3$ ft.; 1000 ft., $-\frac{1}{13}^7$ ft.

Thick Lenses.

185. Find the position of the optical center of a convex meniscus, of which the radii are $r=5$, $s=3$, the thickness $nc=1.5$, and the index of refraction 1.5. (Examples, XVII, 2.)

Ans. $f=-3.75$.

186. With the lens of the preceding Exercise find the source of which the optical center is the image formed by the first surface, and the image formed by the second surface, of which the optical center is the source. (Examples, XVII, 3.)

Ans. $u=-2$, $v=-1.2$.

187. Locate the principal foci and principal planes of a convex meniscus, given $r=5$, $s=3$, $nc=1.5$, $n=1.5$. Construct the images (1) of an object standing outside the principal focus; (2) of a virtual object at U; (3) of an object within the principal focus. (Examples, XVII, 5.)

Ans. $X=-20$, $Y=7.2$; $\bar{x}=-8$, $\bar{y}=-4.8$; $\Phi=12$.

188. (*a*) Locate the principal foci and the principal planes of a double convex lens, given $r=-4$, $s=5$, $nc=1.5$, $n=1.5$. (*b*) Also of a lens with the same surfaces, but with $nc=30$, $n=1.5$; in this case the principal planes are outside the foci. An object outside the principal focus has an erect image; an object in the principal plane has a real image erect and of the same height in the other principal plane.

Ans. (*a*) $X=\frac{6}{11}$, $Y=-\frac{10}{17}$; $\bar{x}=\frac{14}{17}$, $\bar{y}=-\frac{18}{17}$. (*b*) $X=-32$, $Y=50$; $\bar{x}=-72$, $\bar{y}=90$.

189. Locate the principal foci and the principal planes of a planoconvex lens, of which the curved surface has the radius $r=-5$, $nc=1.5$, $n=1.5$. In this case $V=\infty$, and the specifications for Y and $\bar{y}$ fail; we can locate X and $\bar{x}$ and use the formula for p, the distance between the principal planes.

Ans. $X=0$, $\bar{x}=10$, the first principal plane is tangent to the convex surface, $p=0.5$, $\Phi=10$.

190. Locate the principal foci and the principal planes of a double concave lens, given $r=5$, $s=-3$, $nc=1.5$, $n=1.5$. *Ans.* $X=-\frac{10}{17}$, $Y=\frac{8}{11}$; $\bar{x}=-\frac{18}{17}$, $\bar{y}=\frac{9}{11}$.

191. Find the formulas for the principal foci and the principal planes of a sphere used as a lens, given its radius r, and the index of refraction n.

Ans. $e=-\dfrac{2r}{n(n-1)}$; $X=-Y=\dfrac{nr}{2(n-1)}$; $\bar{x}=-\bar{y}=\dfrac{nr}{n-1}$; $\Phi=\dfrac{nr}{2(n-1)}$, $p=0$.

The principal planes coincide at the center of the sphere.

192. Locate the principal foci of (*a*) a globe of water, for which $n=\frac{4}{3}$; also of (*b*) a globe of glass, for which $n=1.5$.

Ans. (*a*) The focus is $2r$ from the center; (*b*) The focus is $1.5r$ from the center.

Combination of Lenses.

193. Prove that when two thin lenses of focal lengths F and G are in contact the focal length Φ of the combination is given by the formula $\dfrac{1}{\Phi}=\dfrac{1}{F}+\dfrac{1}{G}$.

194. If a near-sighted eye can see distinctly at a distance of 4 in., what must be the focal length of a lens which will enable it to see distinctly at the distance of 10 in.? We may take these distances as the distances from the lens.

Ans. The lens is concave, with a focal length of $6\frac{2}{3}$ in.

195. If an object 3 mm. high is placed under a convex lens of 2 cm. focal length, and its virtual image brought to the distance of 27 cm. from the lens, what is the height of the image? *Ans.* $h=37.5$ mm.

196. The object lens of an opera glass has a focal length of 5 in., and the eyepiece a focal length of 2 in. (*a*) How far are the lenses apart? (*b*) what is the magnification? *Ans.* (*a*) 3 in.; (*b*) 2.5.

197. Show that if two convex lenses are placed at a distance from each other equal to the sum of their focal lengths an object in the focal plane of one of the lenses will form an inverted image in the focal plane of the other lens. This combination is used as the inverting system in a telescope made for ordinary use.

198. Two thin convex lenses of focal lengths F and G are placed at a distance d apart. Find a formula for the distances K and L of the principal foci of the combination from the lenses of focal length F and G respectively.

Ans. $\dfrac{1}{K}=-\left(\dfrac{1}{F}+\dfrac{1}{G-d}\right)$, $\dfrac{1}{L}=\left(\dfrac{1}{G}+\dfrac{1}{F-d}\right)$.

199. A thin convex lens of focal length F stands in front of a thin concave lens of focal length $-G$ at a distance d from it. Find the formula for the distance L of the principal focus of the combination from the concave lens when the parallel light is received by the convex lens. *Ans.* $\frac{1}{L} = \frac{1}{G} - \frac{1}{F-d}$.

200. Apply the formula of the last Exercise to the case of a convex lens of focal length $F=10$, and a concave lens of focal length $G=-2$, separated by a distance $d=8.5$. This is essentially the combination used in the telephotographic lens system. *Ans.* $L=6$.

201. Find the ratio of the height of an image cast on the focal plane of the combination of the last Exercise with the height of the image of the same object on the focal plane of the convex lens. *Ans.* 4.

NOTE. — In the following exercises use the formulas of Examples, XVIII, 1.

202. Find the principal foci and the principal planes of a combination of two double convex lenses; given $U=3$, $m=1$, $V=-2$, $n=1$, $d=10$. This is the case represented in Fig. 190.

Ans. $e=5$; $X=-1.8$, $Y=0.8$; $\bar{x}=-3$, $\bar{y}=2$; $\Phi=-1.2$; $p=22$.

203. Show that the construction of Examples, XVIII, 1, fails for the case of two converging lenses so placed that $d=U-V$. This is the telescopic combination; the principal foci and the principal planes are at infinity.

204. Find the principal planes and principal foci of a combination of two thin convex lenses; given $U=6$, $V=-6$, $d=2$.

Ans. $e=-10$; $X=-Y=3.6$; $\bar{x}=-\bar{y}=7.2$; $\Phi=3.6$, $p=-0.4$; the principal planes are reversed.

205. Find the principal planes and principal foci of a combination of two thin convex lenses; given $U=5$, $V=-5$, $d=8$.

Ans. $e=-2$; $X=-Y=12.5$; $\bar{x}=-\bar{y}=25$; $\Phi=12.5$; $p=-32$; the foci and principal planes are reversed.

206. Find the principal foci and the principal planes of a combination of a convex lens, $U=10$, and a concave lens, $V=2$. The distance between them is $d=8.5$.

Ans. $e=0.5$; $X=-200$, $Y=8$; $\bar{x}=-160$, $\bar{y}=-32$; $\Phi=40$; $p=72$. This is essentially the combination used in the telephotographic lens. Notice that the focal length of the combination which determines the height of the image of a distant object cast by it, is increased by the introduction of the concave lens, to four times the focal length of the convex lens, while the focal plane, at Y, is only 4.5 units further back than the focal plane of the convex lens. The combination is used to obtain large pictures of distant objects without undue lengthening of the camera.

Interference.

207. In Young's experiment the distance between the openings is 2 mm.; the distance of the openings from the screen is 1 meter. Yellow light is used and it is observed that the distance between the fifth black band on one side of the central bright band to the fifth black band on the other side is 5.3 mm. Calculate the wave-length of the light. *Ans.* $\lambda=0.00059$ mm.

208. In Newton's experiment to show the rings with yellow light, if the radius of the spherical lens surface is 6 meters, what will be the radius of the fifth black ring? *Ans.* $r=4.2$ mm.

209. Calculate the length of the first half-period element of orange-red light ($\lambda=0.0006$ mm.) at the distance of 1 meter from the center P.

Ans. 0.775 mm.

210. Plane waves of light fall on a slit 0.4 mm. wide, and the 10th diffraction band produced by it on a screen at 2 meters distance is 25 mm. from the center of the pattern. Calculate the wave-length of the light. *Ans.* 0.0005 mm.

211. There are 800 lines to the centimeter in a certain diffraction grating. The red light from hydrogen, corresponding to Fraunhofer's line C, falls perpendicularly upon it. What will be the inclination of its diffraction band or line of the first order? *Ans.* 31° 40′.

212. If a diffraction grating has 600 lines to the millimeter, and the diffraction band or line of the second order is inclined at the angle $\theta=44°\ 50'\ 20''$, what is the wave-length of the light producing it? *Ans.* 0.0005876 mm.

Magnetism.

213. Two equal magnetic poles, 6 cm. apart, exert on each other a force of 196 dynes. What is the strength of the poles?

Ans. 84 units.

214. If one of the magnets of the last Exercise is 8 cm. long, what is its magnetic moment? *Ans.* 672.

215. Find the magnetic intensity at a point on the line of the axis, 100 cm. from the center of the magnet of the last Exercise, (*a*) by the formula of Examples, XX, 1, (*b*) by direct calculation of the forces exerted by the poles. Also find (*c*) the intensity by direct calculation for the same point for a magnet having the same moment 672 and of length 2 cm.

Ans. (*a*) 13.44×10^{-4}. (*b*) 13.91×10^{-4}. (*c*) $13.44+\times10^{-4}$.

216. A magnet is used to deflect an auxiliary magnet in the first position of Gauss. The distance between their centers is 30 cm., and the angle of deflection is 14°. The same magnet executes 100 double vibrations in the earth's field in 296 seconds. Its moment of inertia is 30. Find the value of the horizontal intensity H, and the magnetic moment M of the magnet.

Ans. $H=0.2004$, $M=674.5$.

217. Find the formula by which we may calculate the moment of inertia I of a magnet that is suspended so as to oscillate in the horizontal plane, by observations of its periods of oscillation T_1 when not loaded and T_2 when loaded with another nonmagnetic body, whose moment of inertia K about the axis can be calculated.

Ans. $I=\dfrac{KT_1^2}{T_2^2-T_1^2}$.

218. How much will the magnetic force at a point in vacuum be altered by filling the space with water?

Ans. It will diminish in the ratio of 0.9999895 to 1.

219. Calculate the magnetic induction in a bar of bismuth placed in a magnetic field of intensity 200 with its length parallel to the lines of force. The susceptibility of bismuth is -14×10^{-6}. *Ans.* $B=199.96$.

220. If the strength H of the magnetic field inside a core of iron surrounded by a coil of wire carrying an electric current is 2.67, what will be the magnetic induction in the iron if its susceptibility is 110? *Ans.* $B=3692$.

Static Electricity.

221. The force between two electric charges, one of which is double the other, is measured by a torsion balance or otherwise when the charges are 8 cm. apart, and found to be equal to the weight of 90 mg. What are the magnitudes of the charges?

Ans. 5.31, 10.62.

222. What work will be done in carrying a unit charge to earth from a point at one corner of a square whose side is 10 cm. long, when positive charges of 20 units and 30 units are situated at the two nearest corners and a negative charge of -28.2 units is situated at the end of the diagonal. *Ans.* 3 ergs.

223. Find the number of volts equivalent to the electrostatic unit of potential. *Ans.* 300.

224. A spherical conductor 10 cm. in diameter is at a potential of 6000 volts. Find (*a*) its charge, (*b*) the electric intensity at a point just outside its surface, (*c*) the electric pressure at its surface, (*d*) the energy of its charge.

Ans. (*a*) 100. (*b*) 4 dynes. (*c*) $2/\pi$ dynes/cm.^{2} (*d*) 1000 ergs.

225. What is the capacity of a spherical condenser of which the radius of the inner sphere is 9 cm.; (*a*) when the radius of the outer sphere is 10 cm. (*b*) when it is 9.1 cm.? *Ans.* (*a*) 90. (*b*) 819.

226. What is the capacity of the condenser described in the last Exercise, if the interspace is filled with shellac. *Ans.* (*a*) 315. (*b*) 2866.5.

227. What is the energy of the charge on a spherical condenser of which the inner radius is 10 cm., the outer radius is 10.5 cm., and which is raised to a potential of 900 volts? *Ans.* 945 ergs.

228. (*a*) How is the energy of this condenser altered if the charge is retained on the inner sphere, and aniline is run into the interspace? (*b*) What will the energy become if, instead of the charge, the potential is kept the same?

Ans. (*a*) It becomes 126 ergs. (*b*) 7087.5 ergs.

229. A circular disk, 4 cm. in diameter, is placed in a slightly larger circular opening cut in a large plate (the guard ring), and stands 1.5 mm. from a large parallel plate. With the guard ring used in this way we may neglect charges on the edges of the disk, and consider the lines of force between it and the opposite plate to be parallel. Find the charge on the disk when its potential is 600 volts, and the opposite plate is joined to earth. *Ans.* 13.3 units.

230. Find the electric intensity in the region between the plates. *Ans.* 13.3 dynes.

231. Find the force exerted on the disk by the attraction of the charge on the opposite plate. *Ans.* 88.89 dynes.

232. Find the potential of the disk if the force between it and the plate is 0.6 grams-weight. *Ans.* 1543 volts.

233. The potential of the disk in the arrangement described in Ex. 229 is kept constant; when the opposite plate is at zero potential and at an unknown distance from the disk the force on the disk is 0.6 grams-weight; the plate is then raised to another potential and the plate moved toward the disk through a distance of 0.2 mm. to make the force on the disk again equal to 0.6 grams-weight. Find the potential of the plate. *Ans.* 205.8 volts.

234. Two spheres, one of radius 20 cm. and at a potential of 10,000 volts, the other of radius 10 cm. and at a potential of 4000 volts, are joined by a fine wire. Find the resulting potential. *Ans.* 8000 volts.

235. Find the area of one surface confronting another at the distance of 0.2 mm. and separated from it by a sheet of paraffine, which forms a condenser with a capacity of $\frac{1}{10}$th a microfarad (the millionth of a farad: the farad being the capacity of a condenser on which a charge of one coulomb will produce a potential of one volt). *Ans.* 9835 sq. cm.

The Electric Current.

236. A tangent galvanometer with 60 turns of wire and a radius of 12 cm. is set up in a magnetic field in which the horizontal intensity is 0.2. Find the current in amperes which will produce a deflection of 45° (the galvanometer constant). *Ans.* $1/5\pi$ ampere.

237. Find the formula for the magnetic force due to a circular current of strength i and radius r at a point on the line perpendicular to the plane of the circle through its center and at the distance S from it. *Ans.* $2\pi ir^2/(r^2+S^2)^{\frac{3}{2}}$.

238. A tangent galvanometer is joined in circuit with an electrolytic cell in which silver is deposited by a constant current. In half an hour 671 mg. of silver are deposited. The deflection of the galvanometer is 55°. Find the galvanometer constant. *Ans.* 2.335.

239. How many cubic centimeters of dry hydrogen will be produced by a current of one ampere maintained for 30 minutes? *Ans.* 208 c.c.

240. Calculate the ionic charge in electrostatic units, by the help of the result of the last Exercise. *Ans.* 3.2×10^{-10}.

241. How much copper will be deposited from a solution of sulphate of copper by a current that deposits 800 mg. of silver? *Ans.* 235.4 mg.

242. Three storage cells in series, of negligible resistance, are joined in circuit with a resistance of 120 ohms. What is the current in the circuit? *Ans.* 0.055 ampere.

243. A current of 0.6 ampere traverses a divided circuit of two branches, of which the resistances are 20 ohms and 100 ohms respectively. Find the currents in the branches. *Ans.* 0.5 ampere, 0.1 ampere.

244. What is the single resistance equivalent to the resistances of the divided circuit of the last Exercise? *Ans.* $16\frac{2}{3}$ ohms.

245. If one resistance in a divided circuit is 1000 ohms, and the other is made successively 1, 2, and 3 ohms by sliding one end of the high resistance along the low resistance, show by direct calculation that the currents in the high resistance will be closely proportional to 1, 2, and 3, if the electromotive force in the circuit is constant and if the resistance of the circuit when the high resistance branch is not joined in is 10 ohms.

246. What will be the current in amperes in a circuit of 80 ohms resistance (*a*) when 1 Daniell's cell, (*b*) when 10 Daniell's cells, and (*c*) when 25 Daniell's cells are in the circuit in series, if the resistance of one cell is 0.6 ohm? *Ans.* (*a*) 0.0134. (*b*) 0.126. (*c*) 0.2842.

247. What will be the current in the same circuit (*a*) when 10 Daniell's cells, and (*b*) when 25 Daniell's cells are in the circuit in parallel?
Ans. (*a*) 0.01349. (*b*) 0.01350.

248. What will be the current in amperes in a short, thick wire of $\frac{1}{100}$ ohm resistance if (*a*) 1 Daniell's cell, and (*b*) 25 Daniell's cells in series are in the circuit? *Ans.* (*a*) 1.77. (*b*) 1.799.

249. What will be the current in the same circuit of $\frac{1}{100}$ ohm resistance if 25 Daniell's cells in parallel are in the circuit? *Ans.* 31.76.

250. What will be the heat developed in a minute in a wire of 80 ohms resistance by a current of 0.2842 ampere? *Ans.* 92+ therms.

251. How much heat will be developed in a second in a wire of $\frac{1}{100}$ ohm resistance by a current of 31.76 amperes? *Ans.* 2.4 therms.

252. Show that if the wire in which the heat is thus developed is of platinum with a length of 1 cm. and a cross section of $\frac{1}{10}$ square millimeter, its temperature will rise in one second to nearly 3600° C.

253. Find the quantity of electricity in coulombs induced in a circular coil of 50 cm. radius and with a resistance of 0.1 ohm per meter, by 10 complete rotations in a magnetic field of strength 0.4. *Ans.* 4×10^{-3}.

254. What is the velocity with which the sliding transverse conductor must move along the rails of Examples, XXVII, 8, in a field of unit strength, to produce unit current in the circuit, when the resistance of the sliding conductor per unit of length is one ohm? *Ans.* 10^9 cm./sec.

TABLES OF PHYSICAL CONSTANTS

The constants in these tables are given with sufficient accuracy to serve for use in simple numerical exercises and in elementary laboratory practice.

I. Relation of U. S. Units to Metric Units.

1. 1 inch = 2.54 centimeters.
   1 foot = 0.3048 meter.
   1 mile = 1.60935 kilometers.
   1 meter = 39.37 inches.
   1 meter = 3.2808 feet.
   1 kilometer = 0.62137 mile.

2. 1 grain = 0.0648 gram.
   1 pound = 0.4536 kilogram.
   1 gram = 15.43 grains.
   1 kilogram = 2.2046 pounds.

1 kilogram (kg.) = 1000 grams (gr.). 1 gram (gr.) = 1000 milligrams (mg.). 1 meter (m.) = 100 centimeters (cm.) = 1000 millimeters (mm.). 1 tonne = 1000 kilograms (kg.). 1 ton = 2000 pounds (lbs.).

II. Acceleration of Gravity.

1. For sea level and different latitudes.

$$g_\phi = g_{45}(1 - 0.002662 \cos 2\phi) \text{ where } \phi \text{ is the latitude,}$$

$$g_{45} = 980.6 \frac{\text{cm.}}{\text{sec}^2.} = 32.172 \frac{\text{ft.}}{\text{sec}^2.}.$$

2. For elevation h above sea level, approximately,

$$\text{per 100 feet subtract } 0.00588 \frac{\text{cm.}}{\text{sec}^2.} \quad \text{or} \quad 0.000193 \frac{\text{ft.}}{\text{sec}^2.}.$$

III. Coefficients of Static Friction.

Wood on wood, dry...	0.25–0.50	Metals on metals, dry..	0.15–0.20
Wood on stone.....	0.40	Metals on metals, wet .	0.3
Metals on oak, dry...	0.50–0.60	Iron on stone'.....	0.30–0.70

IV. Relative Densities or Specific Gravities, Approximate.

Aluminium.....	2.6	Crown glass.....	2.5
Copper.....	8.9	Quartz.....	2.65
Iron.....	7.7	Pine wood.....	0.35–0.50
Platinum.....	21.5	Oak.....	0.6–0.9
Silver.....	10.5	Ethyl alcohol.....	0.791
Zinc.....	7.15	Glycerine.....	1.26
Brass.....	8.5	Mercury.....	13.596

Air.....	0.001293	Hydrogen.....	0.000090
Carbon dioxide.....	0.001974	Nitrogen.....	0.001257
Chlorine.....	0.003133	Oxygen.....	0.001430

V. Elasticity.

1. Modulus of volume elasticity, in grams per square centimeter, at 0° C.

Mercury	2629×10^5	Water	206×10^5
Steel	1518×10^5	Glass	405×10^5

2. Rigidity and Young's modulus in grams per square centimeter.

	Rigidity.	Young's modulus.
Aluminium	$241 - 335 \times 10^6$	
Brass	$320 - 410 \times 10^6$	$1000 - 1200 \times 10^6$
Iron	$700 - 800 \times 10^6$	$1700 - 2100 \times 10^6$
Steel		$1900 - 2100 \times 10^6$
Glass	273×10^6	$600 - 800 \times 10^6$
Wood	$7 - 12 \times 10^6$	$70 - 154 \times 10^6$

VI. Surface Tension, in Dynes per Centimeter, at 20° C.

Ethyl alcohol	21.7	Water	72.8
Chloroform	28.3	Mercury	470

Range of molecular action, $50\text{–}118 \times 10^{-7}$ cm.

VII. Gases.

Number of molecules in cubic centimeter at 0° C. and under atmospheric pressure, 4×10^{19}.

Mass of hydrogen molecule, 2.2×10^{-24} grams.

Velocity of mean square of hydrogen, 184,200 cm./sec.

Radius of hydrogen molecules, supposed spherical, 10^{-8} cm.

Number of collisions per second in cubic centimeter of hydrogen, 2.4×10^{29}.

Mean free path in hydrogen, 1.4×10^{-5} cm.

VIII. Density of Water at t° C.

Temp. C.	Density.	Temp. C.	Density.
0	0.999871	30	0.99577
4	1.000000	40	0.99235
10	0.999747	50	0.98820
15	0.999160	75	0.97498
20	0.998259	100	0.95865
25	0.997120		

IX. Density of Mercury at t° C.

Temp. C.	Density.	Temp. C.	Density.
0	13.5956	40	13.4974
10	13.5709	90	13.3764
20	13.5463	140	13.2569
30	13.5218	240	13.0210

X. Coefficients of Linear Expansion, at 40° C.

Aluminium	2.313×10^{-5}	Silver	1.921×10^{-5}
Copper	1.678×10^{-5}	Zinc	2.918×10^{-5}
Iron, cast	1.061×10^{-5}	Glass	0.833×10^{-5}
Platinum	0.899×10^{-5}	Vulcanite	6.36×10^{-5}

XI. Specific Heats at Moderate Temperatures.

Solids.

Aluminium	0.214	Silver	0.056
Copper	0.090	Zinc	0.092
Iron	0.11	Brass	0.088
Lead	0.030	Ice	0.505
Platinum	0.0323		

Liquids.

Ethyl alcohol	0.5475	Kerosene	0.4573
Aniline	0.5155	Mercury	0.0334

Gases under Constant Pressure.

Air	0.23788	Hydrogen	3.406
Water vapor	0.4296	Carbon dioxide	0.2012

XII. 1. Melting Points. 2. Boiling Points. 3. Heats of Fusion. 4. Heats of Vaporization.

	1	2	3	4
Ammonia	.....	−38.5	.....	295 (8°)
Carbon dioxide	.....	−79.1	.....	57.5 (0°)
Chlorine	−102	−33.6	.....	.....
Copper	1100	.....	.....	.....
Iron	1635	2450	23	.....
Lead	326	1450−1600	5.86	.....
Mercury	− 39	357	2.82	62 (350°)
Nitrogen	−208	−194.4	.....	.....
Oxygen	.....	−183	.....	.....
Platinum	1900	.....	.....	.....
Zinc	415	958	28.13	.....
Water	0	100	79.3	537 (100°)

XIII. Vapor Pressure of Water at Various Temperatures, in Millimeters of Mercury.

Temp. C.	Pressure.	Temp. C.	Pressure.
0	4.60	75	288.52
10	9.17	100	760.
20	17.39	150	3581.2
30	31.55	200	11689.0
50	91.98	225	19097.0

XIV. Critical Temperatures θ, Pressures π in Atmospheres, Volumes ϕ, Density δ.

	θ	π	ϕ	δ
Air	−140	39	.....	.....
Ammonia	130	115	.....	.....
Carbon dioxide	30.9	77	0.0047	.....
Ether	19.7	35.77	0.01584	0.208
Hydrogen	−220	20	.....	.....
Nitrogen	−146	35	.....	0.44
Oxygen	−118	50	.....	0.6044
Water	358	.....	0.001874	0.429

XV. Index of Refraction for *D* Lines, Approximate.

Crown glass	1.52	Canada balsam	1.528
Flint glass	1.62	Ebonite	1.6
Rock salt	1.544	Water	1.3336
Fluor spar	1.434	Carbon disulphide	1.64
Diamond	2.47	Air	1.00029
Iceland spar		1.658 ordinary index.	
		1.486 extraordinary index.	
Quartz		1.544 ordinary index.	
		1.553 extraordinary index.	

XVI. Wave Lengths of Light in Centimeters $\times 10^8$ with the Corresponding Element.

B	6870 Oxygen.	E_1	5270 Iron, calcium.
C	6563 Hydrogen.	b_1	5184 Magnesium.
D_1	5896 Sodium.	F	4861 Hydrogen.
D_2	5890 Sodium.	G	4308 Iron, calcium.
D_3	5876 Helium.	H	3968 Calcium.

XVII. Magnetic Susceptibility.

Water	-0.837×10^{-6}	Oxygen at 1 atmo...	0.117×10^{-6}
Carbon disulphide	-0.816×10^{-6}	Air at 1 atmo.....	0.024×10^{-6}

XVIII. Dielectric Constant, Approximate.

Ebonite	2.7	Sulphur	3.0
Crown glass	7.0	Aniline	7.5
Paraffin	2.3	Castor oil	4.7
Shellac	3.5	Petroleum	2.1

XIX. Contact Difference of Potential, in Volts.

	Carbon.	Copper.	Platinum.	Zinc.	Zinc amalgam.
Carbon	0	0.370	0.113	1.096	1.208
Copper	−0.370	0	−0.238	0.750	0.894
Platinum	−0.113	0.238	0	0.981	1.125
Zinc	−1.096	−0.750	−0.981	0	0.144
Zinc amalgam	−1.208	−0.894	−1.125	−0.144	0

XX. Atomic Weights, Referred to Hydrogen = 1, Valencies, Electrochemical Equivalents in Milligrams per Coulomb.

Element.	Symbol.	Atomic weight.	Valence.	Electrochemical equivalent.
Aluminium	Al	26.9	3	0.0936
Carbon	C	11.9	4	0.0311
Chlorine	Cl	35.18	1	0.3672
Copper	Cu	63.1	1	0.6588
			2	0.3290
Helium	He	4		
Hydrogen	H	1	1	0.0104
Iron	Fe	55.5	2	0.2895
			3	0.1930
Mercury	Hg	198.5	1	2.0717
Nitrogen	N	13.93	3	0.0485
Oxygen	O	15.88	2	0.0829
Platinum	Pt	193.3	2	1.0098
Silver	Ag	107.12	1	1.1180
Thorium	Th	230.8	2	1.2042
Uranium	U	236.7	2	1.2353
Zinc	Zn	64.9	2	0.3387

XXI. Electromotive Force of Voltaic Cells, in Volts.

Bunsen	1.86	Leclanche	1.46
Chromate	2.00	Edison-Lalande	0.70
Daniell	1.08	Zn–H_2O–Cu	0.98
Grove	1.80	Clark	1.434 (15°C.)
Storage	2.2	Weston, normal	1.0183 (20°C.)

XXII. Resistivity in Millionths of an Ohm per Centimeter of a Bar One Square Centimeter in Cross Section, at 0° C.

Aluminium	2.9–4.5	Silver	1.5–1.7
Copper	1.58–2.2	Zinc	5.56–6.04
Iron	9.7–12.	Phosphor-bronze	8.5
Platinum	9.0–15.5	Carbon	4092
Mercury	94.	German silver	20.9

INDEX

The numbers refer to pages. Proper names are in italics.

Abbe, 348.
Aberration of light, 350, 392.
Aberration, spherical, 335; chromatic, 347.
Abraham, 539.
Absorption of light, 388.
Accademia del Cimento, 246, 249, 250, 267.
Acceleration, 54; unit of, 55; constant, 58; of falling bodies, 59; angular, 106.
Accelerations, composition of, 62.
Achromatism, 347.
Action, 81.
Adiabatic, 293, 301.
Æpinus, 436.
Air, weight of, 188.
Airy, 165.
Alternating electric current, 486, 518.
Amontons, 269.
Ampère, 191, 192, 468, 469, 470, 474.
Ampere, 473, 510; silver deposited by, 493.
Andrews, 278, 280.
Angle, unit of, 7.
Angular displacement, 105.
  velocity, 105.
  acceleration, 106.
  momentum, 109.
  harmonic motion, 76.
  velocities, composition of, 123.
Anode discharge, 531.
Antinode, 233.
Arago, 376, 382, 383, 392, 485.
Archimedes, 12, 130, 135.
Archimedes' principle, 135, 189.
Aristotle, 40, 188, 211, 309.
Arrhenius, 210, 491.
Astatic needle, 475.
Atmosphere, pressure of, 141.
Atom, 192, 193.
Atomic heat, 262.
Atomic weight, 488.
Atwood's machine, 57, 80, 118.
Avogadro, 191, 192.
Avogadro's law, 191, 198.

Bacon, Lord, 284, 289.
Balance, 28.
Balmer, 390.
Barometer, 139.
Bartholinus, 353.
Beats, 241.
Becquerel, 535.
Becquerel rays, 535.
Bernoulli, Daniel, 143, 193, 219.
Bernoulli, John, 33.
Bernoulli's theorem, 143.
Biot, 264, 466.
Biot and Savart's law, 468, 482.
Black, 189, 254, 256, 258, 259.
Body, 96.
Boiling, 276.
Bolometer, 391.
Boltwood, 538.
Boltzmann, 200, 304, 525.
Borda, 114, 138.
Boscovich, 356.
Boyle, 189, 250, 258.
Boyle's law, 189, 197, 271.
Brace, 385.
Bradley, 349.
Brewster, 378.
Brown, 199.
Brownian movements, 199.
Bubbles, 184.
Bucherer, 539.
Bunsen, 260, 386.

Caignard de la Tour, 277.
Cailletet, 282.
Caloric, 284.
Calorie, 254.
Calorimeters, 259.
Camera, 339.
Campbell, 539.
Capillarity, 176.
Capillary tubes, 183.
Carlini, 158.
Carlisle, 464.
Carnot, 292, 297, 306.
Carnot's theorem, 297, 305, 306.
Cathode discharge, 528, 530; constituents of, 530.
Cavendish, 149, 158, 248, 289, 436, 450.
Cavendish's experiment, 154
Celsius, 248.
Center of gravity, 24.

Center of mass, 97; velocity of, 98; acceleration of, 99.
Centimeter, 6.
Centrifugal force, 86.
Centripetal force, 85.
Charles, 191.
Chemical equivalent, 489.
Chemistry, 193.
Chladni, 240.
Christiansen, 389.
Circle, uniform motion in, 71.
Clausius, 289, 297, 299, 301, 302, 307, 491.
Cohesion, 166, 176.
Collision of bodies, 102.
Colloid, 208.
Color, 388.
Combining volumes, law of, 191.
Common light, 377.
Compass, 396.
Compression, 167, 171.
Conduction of heat, 264.
Conductivity for heat, 264; surface, 264.
Conservation of energy, 287.
Contact, angle of, 182, 186.
Continuity, condition of, 142.
Continuity of liquid and gaseous states, 280.
Coulomb, 203, 399, 437.
Coulomb, 438, 474.
Counter-electromotive force, 509.
Couple, 21; impulsive, 110.
  effect of, on a free body, 110.
Critical angle, 321.
Critical temperature, 277, 280.
Crookes, 198, 528.
Crystalloid, 208.
Cunæus, 431.
Curie, 535, 537, 538.
Curie, Mme., 535, 537.
Cycle, 291, 296.
  Carnot's, 296, 308.
Cycloid, motion in vertical, 66, 74, 75.

Dalencé, 247.
Dalton, 270, 272, 275.
Dalton's law, 275.
Darwin, 272.
Davy, 285, 289, 465, 487.
De la Rive, 523.
De Luc, 256, 267.
Density, 137.
Descartes, 90, 319, 346, 348.
Dewar, 283.
Diamagnetism, 419.
Dielectric, 450; energy in, 456.
Dielectric constant, 451.
Diffraction of light, 353, 365; at an edge, 367; by a slit, 368; colors produced by, 368; theorems concerning, 373–375.
Diffraction grating, 369, 375.
Diffusion of liquids, 207; through membranes, 208; of gases, 210.
Dimensions, 7.
Dispersion of light, 347; anomalous, 389.
Displacement, 62; angular, 105.
Dissociation, 210, 491.
Distance, estimate of, 311.
Dollond, 347.
Double refraction of light, 353, 361, 381.
Drops, 184.
Dufay, 430, 436, 446.
Dufour, 276, 281.
Duhamel, 244.
Dulong, 261, 389.
Dulong and Petit's law, 261.
Dutrochet, 208.
Dynamics, 11, 53.
Dynamo machine, 486.
Dyne, 83.

Earnshaw, 222.
Earth, density of the, 157.
  magnetic properties of the, 398, 402.
Ebullition, 276.
Echo, 225.
Efficiency, 297.
Elastic fatigue, 173.
Elasticity, 166; of volume, 168; of traction, 168; of torsion, 169; of flexure, 170.
Electric arc, 465.
  attractions and repulsions, 429, 430.
  capacity, 448; theorems connected with, 458–459.
  charge, 438; unit of, 438; electromagnetic unit of, 473.
  condenser, 449; discharge of, 498.
  conductors, 429.
  convection, magnetic effect of, 525.
  current, 464; heating action of, 464, 508; chemical action of, 464; magnetic field of, 466; law of magnetic force of, 468; equivalent to magnetic shell, 470; potential of, 472; theories of, 516.
  current, electrostatic unit of, 499, 503; electromagnetic unit of, 473, 502, 503, 510.
  current, induced, 483; theorems concerning, 512–516.
  currents, forces between, 469, 481; energy of, 478–481.
  density, 435, 445.
  discharge, 525–530.
  distribution, 435.
  field, 439, 452; bodies in, 457.

Electric force, law of, 436; lines of, 439; theorems concerning, 444–446; physical lines of, 451; tubes of (*see* Tubes of electric force), 451; flux of, 452.
induction, 432.
induction, 454, 460.
intensity (strength of field, electric force), 439.
machine, 431, 443.
potential, 440; zero of, 443; unit of, 443; contact difference of, 461.
pressure, 435, 446.
resistance, 500; electrostatic unit of, 502; electromagnetic unit of, 503, 510; theorems concerning, 506–508.
spark, 431, 527.
Electricity, 429; theories of, 436, 458; equality of two kinds of, 446; identity of, from different sources, 497.
Electricity, quantity of, 438; unit of, 438; electromagnetic unit of, 473.
Electrification, by friction, 429; by induction, 432, 439.
Electrified conductors, energy of, 446.
Electrolysis (chemical action of the electric current), 464, 487; nomenclature of, 487; laws of, 488; theories of, 490–492.
Electromagnet, 475.
Electromagnetic induction, 483–486.
rotations, 477.
theory of light, 522.
waves, 520, 523.
Electrometer, 444.
Electromotive force, 500; electrostatic unit of, 500; electromagnetic unit of, 502, 511.
Electron, 533–535.
Electrophorus, 435, 442.
Electroscope, 434, 444.
Elster, 534.
Emission theory of light, 355.
Energy, kinetic, 89; potential, 93; conservation of mechanical, 94; kinetic, of a system, 101; conservation of, 287; dissipation of, 303.
Engine, Carnot's, 296, 304; thermodynamic, 297.
Entropy, 301, 306, 307.
Equilibrant, 15, 38.
Equilibrium, 12; about an axis, 13; stable, unstable, neutral, 25, 95.
Equipotential surface, 408, 441.
Erg, 89.
Ether, 355, 357, 376, 525.
Evaporation, 273, 276.
Ewers, 531.
Ewing, 423, 425.
Expansion of solids, 267; of liquids, 268; of gases, 269–271.
Eye, 339.

Fahrenheit, 248, 251.
Faraday, 278, 385, 406, 408, 410, 419, 435, 447, 450, 451, 465, 477, 483, 486, 487, 490, 492, 494, 497, 517.
Feddersen, 499, 526.
Fermat, 321.
Fermat's principle of least time, 321.
Fick, 207.
Fizeau, 350, 393.
Flexure, 170.
Floating bodies, 136.
Fluid, 174.
Fluorescence, 391.
Flux of force, magnetic, 410, 411; electric, 452.
Focal plane, 331.
Focus, of spherical mirror, 314; of refracting surface, 324; of lens, 329, 333, 338; of combination of lenses, 345.
Foot-pound, 89.
Force, 9; scale of, 9; units of, 10, 83; direction of, 10; line of, 11; transmissibility of, 11; moment of, 13, 108.
effect of constant, 55; effect of, on a free body, 110.
electric, lines of, 439; physical lines of, 451; tubes of, 452; flux of, 452; unit tubes of, 453.
magnetic, lines of, 406; physical lines of, 408; tubes of, 409; flux of, 410, 411; unit tubes of, 412.
Forced vibrations, 76, 231.
Forces, parallel, 16–22; moments of, 17–19; parallelogram of, 39, 80; composition of, 40, 43; polygon of, 41; triangle of, 41, 47; resolution of, 41.
conservative and nonconservative, 94; internal and external, 100.
Foucault, 125, 350.
Fourier, 219, 264.
Fourier's theorem, 219.
Franklin, 430, 432, 435, 436, 446, 450.
Fraunhofer, 387.
Fraunhofer's lines, 387.
Freezing, 251; change of volume on, 252.
Freezing mixtures, 250, 258.
Frequency, 212; absolute, 243.
Fresnel, 365, 366, 367, 368, 376, 378, 380, 382, 384, 393.
Friction, static, 49, 202; coefficient of, 49; angle of, 50.
kinetic, 126, 202–205; coefficients of, 126, 203; angle of, 126; laws of, 203.
of gases, 204.
of liquids, 203.

Galileo, 31, 53, 55, 59, 61, 63, 65, 77, 117, 139, 188, 202, 246, 342, 348.
Galvani, 461.
Galvanometers, 475.
Gas, 188, 189; ideal, 271; cooling of, by expansion, 272; expansion of, into vacuum, 272.
Gases, laws of, 189–193; kinetic theory of, 193–201; diffusion of, 210; specific heats of, 261; liquefaction of, 277–283.
Gassendi, 222.
Gauss, 177, 333, 399, 405.
Gauss's theorem, 410, 411, 417.
Gay-Lussac, 191, 270, 272.
Gay-Lussac's law, 191, 270, 271.
Geitel, 534.
Geometrical optics, 311.
Gilbert, 396, 398, 429.
Graham, 208, 210.
Gram, 6, 83.
Gravitation, 149–156; force of, 153; law of, 154.
Gravitation constant, 154.
Gravity, acceleration of, 60, 156, 162.
Gravity, center of, 24.
Green, 440.
Grey, 429.
Grimaldi, 353.
Grotthus, 490.
Guericke, von, 188, 189, 429, 431.
Gyration, radius of, 112.
Gyroscope, 121.

Haga, 533.
Harmonics, 230, 240.
Harmony, 242.
Harris, 444.
Heat, 253; unit of, 254; mechanical equivalent of, 286; conduction of, 264; convection of, 265.
Heat, nature of, 284, 291; kinetic theory of, 289.
Heat, latent, of fusion, 257.
latent, of vaporization, 258.
specific, 255.
Heat capacity, 254, 256.
Helmholtz, 145, 242, 288, 509.
Helmholtz's theorem, 508.
Henry, 485, 498.
Herschel, F. W., 386, 390.
Hertz, 522, 523, 531.
Heussler, 419, 426.
Hittorf, 491, 530.
Hooke, 167, 250, 352.
Hooke's law, 167, 173.
Horse power, 89.
Hull, 525.
Huygens, 65, 66, 69, 71, 74, 90, 114, 157, 354, 355, 357, 358.
Huygens' principle, 358.
Hydrodynamics, 141.
Hydrometers, 138.
Hydrostatic balance, 138.
Hydrostatics, 130.

Image, formed by a mirror, 316; formed by refraction at a spherical surface, 326; formed by a lens, 330.
Impulse, 84; work done by, 91; moment of, 109; effect of, on a free body, 110.
Inclined plane, 29, 29–32, 36, 47; motion on, 57, 60, 61; ball rolling down, 117.
Induction, electric, 454, 460.
magnetic, 421.
of electric currents, 483.
Induction coil, 486.
Inertia, moment of, 107; moments of, 111; principal axes of, 121.
Inertia, principle of, 77.
Ingenhouss, 264.
Interference of light, 361–365, 371–373; of polarized light, 376, 378.
Interferometer, 369.
Ion, elementary, 490.
Ionic charge, 489, 534, 535.
Ions, in electrolysis, 487.
Isentropic, 301.
Isothermal, 279, 292.

Jamin, 369.
Jansen, 342.
Jolly, von, 156, 158.
Joly, 260.
Joule, 89, 198, 273, 286, 287, 288, 289, 302, 464, 508.
Joule's law, 508.
Jurin, 183.
Jurin's law, 183, 186.

Kaleidoscope, 319.
Kamerlingh Onnes, 283.
Kater, 116.
Kauffmann, 539.
Kelvin, Lord (*William Thomson*), 146, 273, 297, 303, 306, 421, 422, 444, 494, 499, 505.
Kepler, 149, 319.
Kepler's laws, 149.
Kilogram, 6, 10, 82, 138.
Kilogram-meter, 89.
Kinetic theory of gases, 193–201.
Kirchhoff, 359, 386, 387, 388.
Kirchhoff's law, 388.
Klatt, 392.
Kleist, von, 432.
Koenig, 240, 242.
Kohlrausch, 491, 505.
Krigar-Menzel, 156, 159.
Kundt, 239, 389.